**Protocol: Updating the International Events Database**  
Part of the Monitoring Hydropolitical Tensions Project  
Drafted by: Melissa McCracken   
28 June 2023

1. **Research Objectives of Update**

* Collect news event articles from June 30, 2008, to 1 July 2023.
  + Note: need to end earlier than the present. The librarian mentioned that it takes some time for the articles to post to Nexis Uni, and it varies depending on the source publisher.
  + Note: Potential for developing a method for extracting articles regularly in the future that could then be coded at intervals to keep the database up to date.
* Collect groundwater-specific news event articles from 30 June 2008 to 1 July 2023
  + If funding and interest, we could extend the groundwater-specific search if we want a more extended time sample of articles.
* Code articles to identify trends in conflict and cooperation occurrences
  + Note: This would include the same codes established in the previous iterations of the dataset and newly developed codes to fit new projects.
* Support qualitative case study work on hydropolitical tension indicators

1. **Methods for Data Collection**
   1. *Past Data Collection*

The first search was done in 2000 by Shira Yoffe and colleagues with the TFDD. The data collection utilized several different databases: International Crisis Behavior Project (1918-1988), Conflict and Peace Data Bank (1948-1978), Global Event Data System (1979-1994), Foreign Broadcast Information Service (1978-1995), Lexis-Nexis (1978 – 2000), and World News Connection (1995-1999). The years in parenthesis are the years covered by the database at the time of the study – Table 1 lists the record of results. In addition, events were collected from known sources to the authors, including the case studies available through TFDD. The database was updated two other times in 2005 and 2008. These two updates used one source, Lexis-Nexis Academic Universe. The other datasets used in the 2000 update either were not kept up-to-date through the needed timeframe for both the second and third updates or were no longer accessible. The update search yielded 755 events from Lexis-Nexis Academic for the 2000 to 2008 timeframe; the total number of search results is not available.

Table 1: 2000 Database Search Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Database*** | ***Approx. Years Covered*** | ***Total Records*** | ***Initial Search Results*** | ***Number of Events*** | ***Number of Interactions*** |
| ICB | 1918-1988 | 412 | 412 | 4 | 4 |
| COPDAB | 1948-1978 | 256,373 | 5,300 | 388 | 549 |
| GEDS | 1979-1994 | 82,778 | 9,500 | 144 | 225 |
| TFDD | 1874-2000 | 200 | 126 | 126 | 535 |
| FBIS | 1978-1995 | n/a | 1,817 | 439 | 770 |
| WNC | 1995-1999 | n/a | 9,589 | 3,221 | 629 |
| LEXIS-NEXIS | 1978-present | n/a | 2,745 | 16 | 17 |

* 1. *Proposed Data Collection*

For this update, Lexis-Nexis Academic is no longer available and has been updated to Nexis Uni. The following link lists the sources: [Sources (tufts.edu)](https://advance-lexis-com.ezproxy.library.tufts.edu/sourceselection/?pdmfid=1516831&crid=18a20d2f-5ba0-418d-8ac1-87e1d663a883)

Nexis Uni is an online searchable database with more than 15,000 news, business, and legal sources from LexisNexis. Searching is available in the following languages: English, Dutch, French, German, Italian, Portuguese, Spanish, Russian, Chinese, Afrikaans, Arabic, Bulgarian, Bulgarian (Romanized), Catalan, Croatian, Czech, Danish, Estonia, Finnish, Greek, Hungarian, Icelandic, Indonesian, Japanese, Korean, Latin, Latvian, Lithuanian, Malay, Norwegian, Polish, Romanian, Slovak, Slovenian, Swedish, Turkish, Vietnamese. Searching in other languages is possible; however, this was tested and seemed to return fewer results than English language search terms.

I consulted with librarians at OSU library as well as searched for alternative databases that could be used in conjunction with Nexis Uni; however, no better database was available through the library. The GDELT dataset was explored as a potential option to address the English language bias in Nexis Uni. The GDELT project does provide links to the raw data/articles, and machine coding is used to code the articles. I had difficulty accessing the data as the interface crashed, and it did not email the requested data; given this, it does not seem useful for our purposes. Fatine had better luck accessing the GDELT project, which provided URLs to news articles every 15 minutes. She found some additional value in using GDELT, including that it is based on open-source data, the data collection is dynamic and updated regularly (every 15 minutes), and it is automated on a large scale. It is based on the Google Cloud Platform and requires more technical knowledge to access the data. Nexis Uni appears to have expanded the number of languages compared to what could be found about Lexis Nexis Academic. Nexis Uni also has a North American bias, with more news results from US new sources. Potentially, GDELT could be used to expand searches in other regions to reduce the English and North American source bias.

Nexis Uni can be accessed through the OSU library and Tufts Libraries.

* 1. *Failed Search Process – With Contract with Nexis Lexis - 2020*

A contract was signed with Lexis Nexis for them to complete the search and deliver the data to us. See the signed contract in the folder. We provided this document and the Excel file OSU\_SearchTerms\_Final\_Nexis which contains all the search terms for the 310 basins, plus a groundwater-only search. This also includes the tributary names for each basin as identified by Kyoko. The tributary names were reviewed, and duplicates were removed; common words were noted to be searched for with proper noun capitalization to avoid additional irrelevant hits, for example, Red River vs red river. Nexis will return the data in the form of a JSON or multiple JSON files, that have been noted which search term the data is from. Alireza Mostafizi developed a Python script to parse the JSON files. This created an Excel table with a row for each search return and an HTML file with the full text of the data.

However, the results processed by Lexis-Nexus were not usable. Search results were not consistent with the manual searching of the data through the web interface and broadly did not reflect the search parameters. The data and information from the failed search are in the folder Events Data Collection 2020

# Secondary Search Process –Starting in September 2022

Nexis Uni has updated the features on the News database, now allowing for the download of full documents and exporting Excel tables of full results. For this process, a manual collection of articles will be used with these features. The following outlines the search process and data organization procedure.

Add data and files within the 2022 Search in the Folder: [Events Data Collection 2022](https://tufts.app.box.com/folder/167957842099) on Box.

An internal search term Excel sheet has been created [Events\_SearchTerms\_Final\_20220720](https://tufts.app.box.com/file/986591840191), This document contains all the search terms, and it has two additional columns linking the search terms to the TFDD BCODE and TFDD Basin Name. Further, it identifies the research assistant responsible for that basin search and data collection.

## First-Level Level Search

1. [Access Nexis Uni](https://advance-lexis-com.ezproxy.library.tufts.edu/api/permalink/e023cb68-357e-4f30-aed5-1a3a4570f7c3/?context=1516831) through the Library website.
2. Click on ***Advanced Search***. Select the ***News*** sub-heading.
3. The search should be conducted for each of the 313 Basins, plus groundwater. For all basins, the following search terms should be used. Terms in italic are new or changed from the previous update. Use the Terms in the [Events\_SearchTerms\_Final\_20220720](https://tufts.app.box.com/file/986591840191) for the most up to date search terms.
   1. In the first box:
      1. Water Specific Terms: *\*water\** OR river\* OR lake OR dam OR stream OR tributary OR diversion OR irrigation OR pollution OR water quality OR flood! OR drought! OR channel OR canal OR hydroelect! OR reservoir *OR groundwater OR aquifer OR drought OR recharge OR "water table" OR "bore hole"*
      2. In the drop-down menu on the right of the box, select **Headline and Lead Section**
      3. Note: These search terms are included in the search and meant to narrow the articles relevant to freshwater.
   2. In the second box:
      1. Select **AND**
      2. Cooperation and Conflict Terms: treaty OR agree! OR negotiat! OR resolution OR commission OR secretariat OR joint management OR basin management OR peace OR accord OR "peace accord" OR settle! OR *cooperat!* OR *collaborat!* OR disput! OR conflict! OR disagree! OR sanction! OR war OR troops OR "letter of protest" OR hostility OR "shots fired" OR boycott OR protest! *OR appeal OR intent OR reject OR threat! OR force OR coerce OR assault OR fight OR demand OR disapprove OR diploma! OR statement OR memorandum*
      3. In the drop-down menu on the right of the box, **select Headline and Lead Section**
      4. Note: These search terms are included in the search and are meant to narrow the articles relevant further to cooperation or conflict.
   3. In the third box:
      1. Select **AND**
      2. Basin Names: Paste the basin-specific names for search of interest from the excel sheet. These are the TFDD basin names and tributary names from the list in the [Events\_SearchTerms\_Final\_20220720](https://tufts.app.box.com/file/986591840191).
      3. In the drop-down menu on the right of the box, select **Headline and Lead Section**
   4. In the fourth box:
      1. Select **AND NOT**
      2. Exclusionary Terms: ocean OR navigat! OR nuclear OR "water cannon" OR "light water reactor" OR "mineral water" OR "hold water" OR "cold water" OR "hot water" OR "water canister" OR "water tight" OR " water down" OR "flood of refugees" OR Rivera OR Suez OR Panama OR oil OR drugs OR "three gorges*" OR waterski OR watermelon OR dishwater OR waterproof OR “water resistant” OR “water bath”*
      3. In the drop-down menu on the right of the box, select **Headline and Lead Section**
      4. Note: These AND NOT terms must come last in the order of the search boxes otherwise, Nexis Uni has issues.
      5. Note: These search terms are not included in the search and are meant to exclude events that are not related to freshwater bodies.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 1: Advanced Search – First-Level Search

1. Date Range: Select ***Date is between…*** and select an appropriate date range for a manageable amount of results if the search is being done manually, like 1 year. If able to get full search results (<10,000), then the search should be between 30 June 2008 and 1 July 2023.
   1. As of 28 June 2023, we are shifting the end date from 15 July 2022 to 1 July 2023. For all searches that have not been completed/started will use these dates. One student will be assigned the task of completing the search for 15 July 2022 to 1 July 2023 for the 183 basins where the search has already been completed.
2. Document ward length: this can be left blank.
3. Language: this can be left blank.
4. Source: this can be left blank

Graphical user interface, application

Description automatically generated

Figure 2: Advanced Search – First Level Search Parameters

1. Click on ***Search*** to initiate the search. Once the first level of the search has been completed, the search time span can be narrowed to reduce the results to below 1000 for the excel download or 100 for the pdfs.

## Primary Search: Groundwater

1. **For the groundwater search only**,
   1. In the Third box
      1. Select **AND**
      2. Basin Names: Paste the groundwater search terms from the excel sheet. [Events\_SearchTerms\_Final\_20220720](https://tufts.app.box.com/file/986591840191).
      3. In the drop-down menu on the right of the box, select **Headline and Lead Section**
   2. In the fourth box:
      1. Select **AND NOT**
      2. Exclusionary Terms: ocean OR navigat! OR nuclear OR "water cannon" OR "light water reactor" OR "mineral water" OR "hold water" OR "cold water" OR "hot water" OR "water canister" OR "water tight" OR " water down" OR "flood of refugees" OR Rivera OR Suez OR Panama OR oil OR drugs OR "three gorges*" OR waterski OR watermelon OR dishwater OR waterproof OR “water resistant” OR “water bath”*
      3. In the drop-down menu on the right of the box, select **Headline and Lead Section**
      4. Note: These AND NOT terms must come last in the order of the search boxes otherwise, Nexis Uni has issues.
      5. Note: These search terms are not included in the search and are meant to exclude events that are not related to freshwater bodies.

## Second Level Search

1. At the top of the page, turn ***Group Duplicates*** to ***On***

A picture containing logo

Description automatically generated

Figure 4: Second Level Search – Group Duplicates

1. If needed, the time range can be further limited using the ***Timeline*** Narrow By filter in the left panel. This could be used if the search within results returns more than 10,000 articles, as Nexis Uni will not pull more than that number of results. If the date range is narrowed, be sure to collect data for the full-time span to ensure no gaps in the data.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 5: Example of Second level search results.

* 1. *Data Collection and Organization Procedure*

1. Identify a basin to run the search in the excel sheet.
2. Create a new folder in the [Events Data Collection 2022](https://tufts.app.box.com/folder/167957842099) titled as the BCODE for the basin.

## Create the Results List Excel Table

1. Run the search as described. Once complete:
   1. Click **Sort by → Date (oldest to Newest).**

Graphical user interface, application

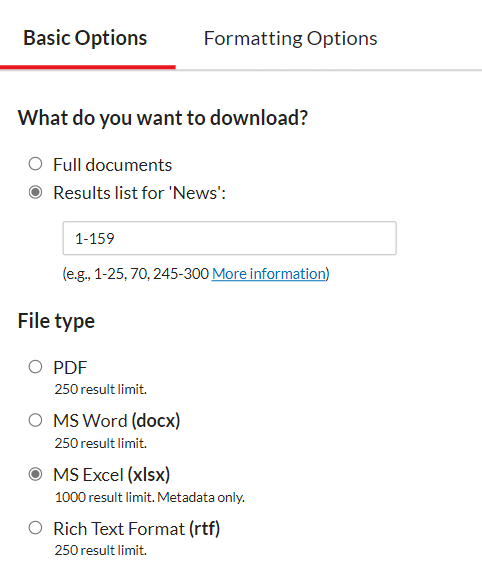
Description automatically generated

* 1. Note how many results are in the search. Then click Download

A picture containing text

Description automatically generated

* 1. Download Results List
     1. Select: Results List for ‘News.’ In the box, type 1-N, where N=the total number of results up to 1000. If there are more than 1000 results, then this download will need to run as many times as needed to collect all the results.
     2. For searches over 1000 results, reduce the time range before downloading and note the date range. Then follow the steps to download. Change the time range to the next range, download the next amount, and repeat.
        1. Nexis Uni will not recognize results values 1000-1999 and above in the ‘results list for news’ box. So, you must use the time feature to reduce it to below 1000.
     3. Select: MS Excel
     4. FileName: Type ResultsList\_BCODE\_202307. Where the BCODE, is the BCODE for the basin the search was run for.



* + 1. Then click download. → this will create a zip file of this name; however, it does not rename the file itself. Extract the zip file, rename it to ResultsList\_BCODE\_202307, and save it in the respective basin folder you created on Box.
    2. If there are more than 1000 results, repeat the download process until all the metadata for the articles has been downloaded.
       1. For searches over 1000 results, reduce the time range using the “narrow by timeline” before downloading and note the date range. Then follow the steps to download. Change the time range to the next range, download the next amount, and repeat.
       2. Nexis Uni will not recognize results values 1000-1999 and above in the ‘results list for news’ box. So, you must use the time feature to reduce it to below 1000.
  1. Updated the Results List
     1. Open the ResultsList\_BCODE\_202307, you just created.
     2. If there were more than 1000 results, combine the data into one Excel sheet.
     3. After the Date Column, add three new headings.
        1. File Name
        2. Page Numbers
        3. Preliminary Review: Does the article meet the inclusion criteria? (Y/N/Needs Review)
     4. Highlight the headers of the table, then from **the Home tab → Editing → Sort and Filter → Filter.** To turn on the filters for the table.
     5. Save the file

## Create the PDF of Full-Text Documents.

1. Make sure that data is still: **Sort by → Date (Oldest to Newest).**

Graphical user interface, application

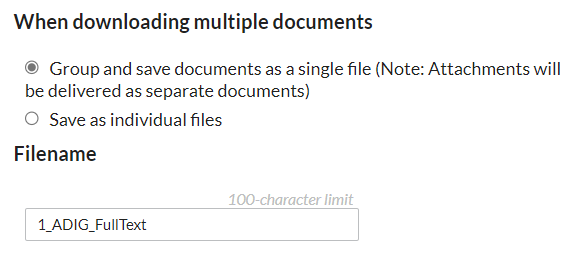
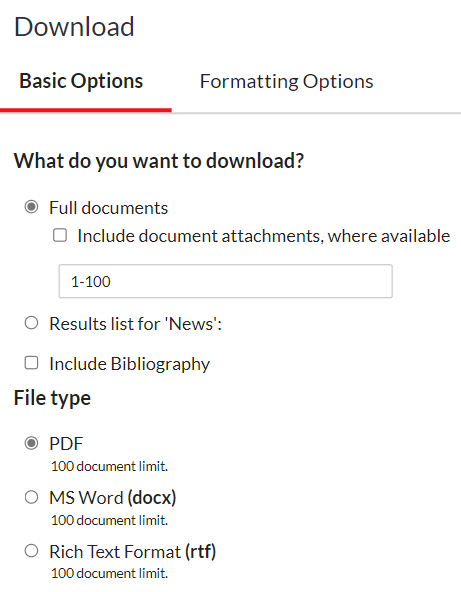
Description automatically generated

* 1. Note how many results are in the search. Then click Download

A picture containing text

Description automatically generated

1. Download Full Documents
   1. Select: Full Documents → In the box, type 1-N. Where N is the number of results up to 100. The interface will only download 100 full-text documents at once. If there are more than 100 results, then the process will need to be completed several times to download the full text of all the articles.



* + 1. For searches over 100 results, reduce the time range using the “narrow by timeline” before downloading and note the date range. Then follow the steps to download. Change the time range to the next range, download the next amount, and repeat.
    2. Nexis Uni will not recognize results values 100+ and above in the ‘Full documents’ box. So, you must use the time feature to reduce it to below 100.
  1. File Type: Select PDF
  2. For the ‘When downloading multiple documents’: check the box - *Group and save documents as a single file*
  3. FileName: Type → 1\_BCODE\_FullText. Use 1 for the first download, 2 for the second (e.g., 2\_BCODE\_FullText), and so forth. BCODE should be the respective BCODE for the basin the search was run for.
  4. Click Download. → Note that this may take some time to create and download.
     1. Note → There is also a chance it will not create the right file name. Please double-check and change the file name as needed.
  5. If there are more than 100 results, repeat the download process until all articles have been downloaded.
     1. For searches over 100 results, reduce the time range using the “narrow by timeline” before downloading and note the date range. Then follow the steps to download. Change the time range to the next range, download the next amount, and repeat.
     2. Nexis Uni will not recognize results values 100+ and above in the ‘Full documents’ box. So, you must use the time feature to reduce it to below 100.
  6. Save all the PDFs in the Basin folder you created on Box.

1. Once all PDFs have been downloaded and the Excel sheet has been created, return to the [Events\_SearchTerms\_Final\_20220720](https://tufts.app.box.com/file/986591840191) and fill out **Y in the Has search been completed and all files downloaded? (Y/N/Issues).** Add any notes that may be relevant or any search issues that arose in the Search Notes Column.
2. Add a note if the basin is over 10,000+.

**Final Steps**

1. Put the filename of the PDF in which the article is located in the filename column.
2. Conduct a preliminary review of the titles/skim the articles to see if the articles meet the inclusion criteria.
   1. Inclusion criteria: deals with water as a consumable resource and is international.
      1. If obviously, yes or no, put Y/N in the Preliminary Review: Does the article meet the inclusion criteria? (Y/N/Needs Review)
      2. If unsure, put Needs Review.
3. Once complete, return to the [Events\_SearchTerms\_Final\_20220720](https://tufts.app.box.com/file/986591840191) and fill out the **Y in Has preliminary review been completed? (Y/N/Issues)** for that basin. Add any notes that may be relevant or any search issues that arose in the Search Notes Column.

## Troubleshooting Errors

Nexis Uni can have a less than user-friendly interface. Here are a few things to keep in mind:

* If it is running slow, try logging in to the university’s VPN.
* For the Excel downloads, the max per session is 1000, and the pdf is 100. The error may or not show up after downloading that number of files.
  + If you get the error that you have reached the max downloads/delivery requests per session, then recommend the following things:
    - Clear the cookies, cache, browsing history, etc., from the browser, then close and reload the browser and try again.
    - Use an incognito window and close and reopen when the error occurs.
    - Log into a different browser to continue.
    - Use a second/different computer.
    - Wait 15 mins to 1 hour and try again.