MEMORANDUM

29 October 2013
To: *Political Analysis*
From: Vito D'Orazio, Steven T. Landis, Glenn Palmer & Philip Schrodt
Re: Lexis Nexis Document Collection

Please note that we cannot provide you with the Lexis Nexis downloads due to copyright and licensing restrictions. However, below the instruction manual we used to retrieve the news reports from Lexis Nexis for MID4.

Using Lexis Nexis Academic Database for MID4 Incident Collection[1]

General Information: For incident collection, we use Lexis Nexis Academic Database. This search engine is available via the Penn State University Libraries website at http://libraries.psu.edu/psul.html. Opening Lexis Nexis Academic brings you to the main page titled "Lexis Nexis Academic: General Search". This page sets you into the *Easy Search* mode for data collection.

I. Using *Easy Search*: The page displays a large blank box for entering text called "search terms" and several small search options to check under the heading "search within", and finally a place to "specify date" of your search. When using *Easy Search,* one must specify the specific news sources that will be used in the "search terms" box in addition to specify the search words that will be used. Properly constructed, this will look like a paragraph of text connected together using connector words and parentheses:

A. **Connectors:** These are words that we use to connect the news sources and search words together in a logical argument for Lexis Nexis Academic to understand. The typical connector words that you will use are OR & AND (capitalized). There are other connector words that can be used in Lexis Nexis. For more information on the visit the help directory under "Search Connectors and Commands".

[1] Lexis Nexis periodically updates the interface of its website. These instructions may require updating from time to time.

B. Publication Specification: To create your search, you must first specify the publications you will use. In order for Lexis Nexis Academic to understand your publication specification you must preface your publication with the following: pub(……..). The letters "pub" tells Lexis Nexis that you want to search the following publications, which are contained in between the parentheses. Each publication needs to be separated by the connector OR (see above). The publication names are somewhat tricky however, and you need to be careful when entering these. For some reason, Lexis Nexis identifies certain terms and not others pertaining to publications. For example, if you want to search for "New York Times" you type either "New York Times or NYT". However, if you type just "Times" Lexis Nexis will search any newspaper with the word "Times" in its name (i.e. London Times, New York Times, St. Petersburg Times etc.). Alternatively, sometimes Lexis Nexis will not recognize a specific name of a news source when you would expect it to. For instance, if you type "FOX news" Lexis Nexis will not recognize it. Other times, if you type the proper name of a publication like "News Straits Times Malaysia" this will also not appear. Finally, if you specify a publication that is contained within an aggregator (see below), Lexis Nexis will not recognize this either. Thus, the point is that when you create your publication list, you need to check the results to see if Lexis Nexis is actually collecting data from the publications you want.

C. Headline Choice: Once you create your publication list you need to create your "Headlines Choice". This MUST be connected to your publication list using the connector AND (see above). Like earlier you need to preface your headline list as the following: hlead(…….). This is creates a logical command that tells Lexis Nexis you want to search for the following headlines in the prior publications section. Each headline needs to be separated using the connector OR (or AND, but this is used less often because it is more restrictive). The headlines we use here are going to be related to MID4 incidents. So words like the following: rocket OR attack OR battle OR fired etc. This headline list is the most comprehensive part of the process. You need to be creative and make sure that you use words that relate to potential incidents. Typically, this headline string will be created by a senior person in charge of the

collection process and remains consistent through the entirety of the project.[2] Within the headline list, you can use certain "Wildcard Characters" (see the help section for more information) that tells Lexis Nexis to be slightly more flexible in its search. In most cases the "wildcard character" that you will use is the ! (exclamation point). For instance, type the root of a word followed by ! : fir! Searches words starting with the root fir. So Lexis Nexis will grab the following headlines: fired, fire, firing. But this will also grab headlines like "first". So you need to think carefully about using this "wildcard character" because it may provide you with extra information that you do not need. Finally, once you create your headline list. You need to add a final section specifying certain headlines for Lexis Nexis to exclude certain stories that often clog up and pervade the returns. Connect this section using: AND NOT(……) with words like natural disaster OR sports OR business OR weather etc.

D. Example of a complete Search String in *Easy Search:* The following is an example of complete search string following the directions laid out above:

pub(Xinhua OR The Times London OR The Hindu OR DPA OR Boston Globe) AND hlead(air base OR air strike OR airbase OR aircraft OR airstrike OR alert OR antiaircraft OR armed OR armo! OR arms OR army OR artillery OR attack OR batteries OR battery OR battle OR battleship OR block! OR bomb OR border OR buildup OR carrier OR casualties OR casualty OR cease OR ceasefire OR cease-fire OR clash! OR combat OR conflict OR crisis OR cruiser OR damage OR declare war OR defence OR defense OR defensive measures OR defian! OR deploy! OR destroy OR detained OR dispatch! OR display of force OR dispute! OR embargo OR erupt! OR fight! OR fire OR fired OR forc! OR fortification OR hit OR hostile OR incursion! OR infantry OR interstate OR invasion OR jet OR kill! OR launch! OR liberate OR line of control OR maneuver OR milit! OR missile! OR mobiliz! OR mortar OR naval OR nuclear OR occup! OR offensive OR operation OR patrol! OR peace decleration OR pullback OR radar OR raid!

[2] MID4 uses a systematic headline and publication list. Please consult the Supplementary Materials for further information.

OR recon! OR reinforcement OR reprisal OR retali! OR rocket OR security OR seiz! OR shell! OR shoot OR shot down OR show of force OR shrapnel OR skirmish OR soldier! OR squadron OR stronghold OR subside! OR target OR tension! OR territ! OR threat! OR trade fire OR troop OR truce OR ultimatum OR USS OR vessel OR violat! OR violence OR vows to OR war OR warn! OR warplane OR warship OR weapon! OR weapons OR withdraw!) AND NOT subject(sports OR business OR baseball OR health OR natural disasters)

E. Other Issues in *Easy Search (AND Power Search)*: In easy search, you must also specify the "search within" section to search three areas: Major US and World Publications (English), News Wire Services, TV and Radio Broadcast Transcripts. Finally, for the date you must specific the "date is" choice and go day by day for each month to collect the data. One very important thing to know is that Lexis Nexis only returns 1000 hit results. If you hit 1000 then your search string is too big. To fix this problem you must either reduce the number of publications you are searching within or break up the headline into multiple parts and run each separately for that day in order to collect all of the results.

F. Downloading in the results in *Easy Search (AND Power Search)*: Once you click the search button, your results will appear shortly afterwards on a new screen. To download the files you must click the "Download Documents Icon" located at the top right of the screen, shaped like a floppy disk. Clicking this button opens a new window that allows you to specify how the manner and type of download you want. Create your download to match the following: Format is "text", uncheck all of the "Page Options" section, Font is "Courier" with the other two boxes unchecked. Finally, Lexis Nexis only lets you download 500 at a time. So if your results are 500 or less you click "all documents" option under the Document Range. Otherwise, you must specify 1-500 and download that first. Then repeat the process for the remaining files (i.e. 1-500, then 501-958). Then, click the "download" button. Finally, after a few seconds of formatting this creates a blue hyperlink allowing you to download it directly to a folder, desktop, etc. depending on the right click option you choose.

G. Saving Files Downloaded Files & Organizing Them: When saving and naming the file, use the following format: MONTH_DATE##_YEAR####_LN_NP_#. Use three letter abbreviations for large named months over 5 letters. Us (e two numbers to indicate the date (i.e. 01, 02, 03.....21), Record the year, LN stands for Lexis Nexis, NP stands for news reports, and the number at the end indicates the download number for the day (sometimes you may have 2 or more). See the following example:

Jan04_2001_LN_NP1

Separate these downloads into corresponding month folders titled the by the month and year. Ex. January 2001. Then once you have the whole year collected place the 12 folders into one folder titling the name of the search string used, and your last name and semester that you collected the files. For example:

"BBC Search String for 2003 Incident Collection, Landis, Spring 2010." [3]

H. Lastly: Once all of your search strings are complete, place a copy of them on the Penn State MID Shared drive in an appropriately named folder, with a title that's meaningful Finally, depending on the preferences of the project directors, compress and burn the files onto a DVD; and send them off to be processed via SVM.

II. Using *Power Search* for Specific Sources and Aggregators: Sometimes we want to search only specific news publications or search within a news aggregator (which is a collection of many small newspapers and news outlets) for incidents. In these special cases, it's better to use *Power Search*, picking sources directly and building a specific list. This process is slightly more complicated than using *Easy Search*, but worth the time if you need to collect certain data.

A. Building Your Source List: From the initial Lexis Nexis Academic Search Page, (discussed above in "General Information"), click the "Sources Tab" at the top left of the screen. Next, click the "Find Sources Tab" at the top left of the new screen that appears. Then, click the "Source Name (A-Z)" option located on the top left of the

[3] Note that these titles are dependent on the SVM software commands and may need to be renamed prior to processing.

screen that appears. After you've completed this, you can now search alphabetically through every single source in Lexis Nexis Academic. Select the desired sources and they become automatically compiled and saved into a source list on the right hand side of the screen in a box called "selected sources". Once finished, click the red "ok-continue button" on the top right hand side of the screen. After clicking "ok-finished" you are brought to a screen similar to the *Easy Search* screen.

B. Searching Using a Selective Source List: From the screen, now enter your headline and connectors describe in the *Easy Search* process. The only difference here is that you begin with hlead(…..) and do not need to specify publications. Pick the corresponding date, and follow the same download instructions and save process as described above in the *Easy Search*.

C. Lastly: One thing to keep in mind is that this compiled source list disappears after you close Lexis Nexis. So you need to remember what sources you used. Write them down separately somewhere else or refer to the provided search strings.