THE NUCLEAR LATENCY DATASET: Codebook*

Matthew Fuhrmann, Texas A&M University

With critical assistance from Benjamin Tkach (King University) and Molly Berkemeier (Texas A&M University)

Version 1.2: February 18, 2016

*The NL dataset is part of the "Correlates of Sensitive Technologies" (COST) project, funded by the United States Department of Energy, Nuclear Energy University Programs (NEUP 3149). We thank our collaborators on the project – particularly Garill Coles, Valerie Lewis, and Paul Nelson – for their helpful feedback and assistance on the production of this dataset.

Overview

The Nuclear Latency (NL) dataset provides information on military-relevant nuclear facilities built globally from 1939 to 2012. It includes details about roughly 250 uranium enrichment and plutonium reprocessing (ENR) plants built in more than 30 countries. States in possession of these facilities generally have the latent capacity to make nuclear bombs. ENR plants are therefore particularly sensitive from a nuclear proliferation standpoint.

The NL dataset is described in: Fuhrmann, Mathew and Benjamin Tkach. 2015. "Almost Nuclear: Introducing the Nuclear Latency Dataset," *Conflict Management and Peace Science* 32(4): 443-461. This codebook provides information to help users interpret the data contained in the file "NL Dataset v1.2" and underscores some of the dataset's limitations. All of the files discussed here are available for download at `www.matthewfuhrmann.com/datasets.html`.

It is important to note that we made some updates to the NL dataset after the publication of "Almost Nuclear." Users of the dataset may therefore notice some small changes in the years for which a few countries had ENR plants in operation. The total number of facilities cited in our published paper (241) also differs from the number of plants that appear in the full dataset (253), since we added some facilities based on new information that emerged. In addition, the updated dataset includes one country (Spain) that was excluded from the version described in the above-referenced article. The current version of the dataset is 1.2; we encourage interested individuals to use this updated version rather than the replication files for the analysis carried out in "Almost Nuclear."

We have assembled a country-year dataset that provides two indicators of nuclear latency (see the file "Country-Year Dataset"). The first considers a country to have nuclear latency if it operates at least a laboratory-scale ENR plant in a given year. A second, more restrictive measure only affords a state with latency status if it has a pilot plant in operation. These are by no means the only ways to measure nuclear latency. Researchers may construct various alternative measures using the information contained in the NL dataset, based on their research needs and objectives.

We produced country-level coding sheets that provide the sources we used to assemble the dataset (see the file "Country Coding Sheets"). These coding sheets also explain the rationale behind some of our judgments, particularly those we made when the evidence was ambiguous. Researchers who wish to know more about particular coding decisions may consult these documents for further information.

Unit of Observation

The unit of observation in the the NL dataset is the ENR facility. A site becomes a "facility" once nuclear materials are introduced with the *intent* to enrich uranium or separate plutonium.[1] This definition excludes some R&D activities. Countries may develop or use ENR-related technology without introducing materials for the purposes of uranium enrichment or plutonium reprocessing. Swedish scientists, for example, built centrifuges during the

[1] The dataset includes some plants that were not finished by the end of 2012 – and therefore do not meet this criterion – to capture states' interest in sensitive nuclear technology.

early 1970s at the Royal Institute of Technology in Stockholm.[2] We did not find evidence that they used centrifuges to enrich *uranium*, however, even though the technology could have theoretically been used for this purpose.[3] This case is therefore not included in our dataset. To further illustrate this distinction, consider a second example: the hot cell at Riso National Laboratory in Denmark. Hot cells may be used to separate plutonium – but they are also employed for other purposes, such as the production of radiopharmaceuticals. The Danish facility is excluded from the NL dataset because we did not find evidence that plutonium was separated there.

Variables

For each facility, we created the following variables:

- COUNTRY_NAME. The name of the country in which the nuclear plant operated.
- FACILITY_NAME. The name of the nuclear plant.
- FACILITY_AMBIGUITY. The line between a lab-scale ENR plant, as we define it, and R&D activities is sometimes fuzzy. In several cases we were unsure whether small-scale enrichment or reprocessing occurred at a site, though we suspect that it did. This variable denotes these cases. Some researchers may wish to exclude these facilities from their analyses since the sites may not meet our criterion for what constitutes a facility.
    - 0. No ambiguity.
    - 1. May have been an R&D site without any uranium enrichment or plutonium separation.
    - -77. Facility was not operational by the end of 2012.
- ENR_TYPE. The technology employed at the nuclear plant.
    1. Reprocessing
    2. Gaseous diffusion
    3. Centrifuge
    4. EMIS
    5. Chemical & ion exchange
    6. Aerodynamic isotope separation
    7. Laser
    8. Thermal diffusion

[2] Bonnevier, Bjorn. 1970. "Experimental Evidence of Element and Isotope Separation in a Rotating Plasma," *Plasma Physics* 13: 763-774.

[3] The Swedish experiments achieved separation in a hydrogen-argon mixture (and a hydrogen-deuterium mixture). See Bonnevier (1970, 773).

- SIZE. The capacity of the nuclear plant based in part on descriptions employed by the International Atomic Energy Agency's (IAEA) Integrated Nuclear Fuel Cycle Information System (INFCIS) database.

    1. Laboratory. The IAEA defines lab-scale facilities as follows: "The facility or process line is being operated in a laboratory to examine the applicability of a process." As noted above, ENR-related R&D activities are not always classified as lab-scale plants. One can think of lab-scale ENR facilities as those that enrich uranium or reprocess plutonium on a very small scale.
    2. Pilot. This means, according to the IAEA, that "The facility or process line is being operated as a precursor of a commercial or industrial facility or process line."
    3. Commercial. "The facility or process line is being operated in commercial or industrial scale," based on the IAEA definition.

- CONSTRUCTION_START. The year in which construction on the facility began. Information on construction start dates was unavailable for some facilities; missing values are coded as "-99."

- CONSTRUCTION_END. The year in which construction on the facility ended. A value of "9999" indicates that construction was not finished as of 2012. Missing values are coded as "-99."

- OPERATION_START. The year in which nuclear materials are first introduced into the plant. Materials may be introduced well before a plant successfully produces significant quantities of enriched uranium or plutonium. A value of "9999" indicates that materials were not introduced as of 2012. Missing values are coded as "-99."

- OPERATION_END. The year in which nuclear materials stop being processed in a plant. The end of operation does not imply that a facility no longer physically exists. A value of "9999" indicates that materials were not introduced as of 2012. A value of "7777" indicates that a plant was still operating as of 2012. Missing values are coded as "-99."

- OPERATION2_START. A small number of facilities stop processing nuclear materials but then resume enrichment or reprocessing at a later date. For such cases, this variable indicates the year in which materials were re-introduced into a plant. We code this variable "NA" if there was not a second operational start date.

- OPERATION2_END. The year in which nuclear materials stopped being processed in a plant for the second time, if applicable. We code this variable "NA" if there was not a second operational end date.

- COVERT. A variable indicating whether work on a nuclear plant began in secret. To code a facility as covert, we required evidence that a state engaged in a concerted campaign to conceal a plant's existence. Moreover, the state needed to hide the facility itself – not just its true intentions. A plant would *not* be considered "covert," for example, if it was publicly announced but used for military purposes unbeknownst to domestic or international audiences. This was often a difficult variable to code: sources rarely state things like "this was a covert plant," and the lack of evidence indicating that a facility was publicly announced does not imply that it was, in fact, secret. We encourage researchers to use this variable with caution.

0. Did not begin covertly.
1. Began covertly.

-99. It could not be determined whether a plant began covertly.

- IAEA. A variable indicating whether a nuclear plant was under IAEA safeguards. In some cases, it is difficult to verify whether the IAEA actually inspected a plant. We assume, in those instances, that facilities are under safeguards if a state has a comprehensive safeguards agreement in place with the IAEA *and* the facility is openly developed for non-military purposes. It is important to note that safeguards work differently for nonnuclear states and declared nuclear weapon states (China, France, Russia, the United Kingdom, and the United States). The latter group is *not* required to place nuclear plants under Agency safeguards. However, they may enter into Voluntary Offer Agreements (VOAs) with the IAEA. VOAs list civilian facilities, including enrichment and reprocessing plants, that *could* be inspected by the IAEA. Yet, in practice, facilities in nuclear weapons states are rarely inspected due in part to budget constraints. We code plants as being under safeguards in nuclear weapons states if they are eligible for inspection per a VOA, regardless of whether they are actually inspected.

  0. The plant was *never* under IAEA safeguards.
  1. The plant was subject to IAEA safeguards at some point; this does not imply that the IAEA inspected (or even knew about) the plant for all of the years that it operated.

  -99. It could not be determined whether a plant was under IAEA safeguards.

- REGIONAL_SAFEGUARDS. A variable indicating whether a plant was subject to safeguards instituted by a regional institution (e.g., Euratom).

  0. The plant was *never* under regional safeguards.
  1. The plant was subject to regional safeguards at some point.

  -99. It could not be determined whether a plant was under regional safeguards.

- MILITARY. A variable indicating whether the nuclear plant had military applications. This is the case if one of three conditions are met: (1) sources revealed that the plant was part of a nuclear weapons program, (2) the plant produced or was meant to produce fissile material for bombs, or (3) the military was involved in the construction or operation of the plant. Note that some plants serve military *and* civilian purposes. This variable codes plants as military if they met at least one the above criterion at *any* point in time.

  0. Civilian.

1. Military.

-99. It could not be determined whether a plant had military applications.

- MILITARY_AMBIGUITY. There is often uncertainty surrounding the purpose of a nuclear plant. In some cases it appears that countries are seriously thinking about building a nuclear arsenal but they do not have a dedicated bomb program per se. This would include Algeria, Norway, and Spain, among other countries. We included this variable to denote borderline cases where there may have been some military purpose behind an ENR plant but there is no definitive evidence of a concerted bomb program.

  0. Not ambiguous.
  1. Ambiguous. The plant *may* have had military applications at some point. This coding could also imply that a plant is coded as having military purposes but it may have, in fact, been civilian.

- MULTINATIONAL. A variable indicating whether the facility was owned by entities in more than one country.

  0. Not multinational.
  1. Multinational.

- FOREIGN_ASSISTANCE. A variable indicating whether a plant was built with state-backed foreign aid. Simply importing materials or commodities from foreign firms does not necessarily imply that foreign assistance has occurred, according to our coding rules; the government must have approved the aid.

  0. No foreign assistance.
  1. Some foreign assistance provided.

  -99. It could not be determined whether a plant was built with foreign assistance.

- FOREIGN_ASSISTANCE_AMBIGUITY. It is sometimes unclear whether foreign assistance (as we define it) has occurred. We created a variable to capture potentially ambiguous cases.

  0. Not ambiguous.
  1. Potentially ambiguous.

Caveats and Limitations

The NL dataset is the most comprehensive source for information on the global development of sensitive nuclear facilities that is publicly available. It nonetheless has some key limitations of which users of the dataset should be aware. Four particular challenges we faced in constructing the NL dataset warrant some discussion.

First, it is likely that some facilities are missing from the dataset. We relied exclusively on open source materials. To be sure, there is a remarkable amount of information on countries' ENR-related activities available in the public record. Yet our dataset is inevitably incomplete. Countries often attempt to conceal information about their sensitive nuclear activities, which reduces our confidence that a fully complete picture of global ENR-activities exists in the available literature. We find it plausible, for example, that North Korea had a smaller uranium enrichment plant in operation before the industrial-scale plant at Yongbyon came online around 2010. There is no clear evidence of an earlier plant operating, however, so we do not include one in the dataset. It would also not be surprising if Iran has (or had) additional ENR plants about which we do not know. We took measures to maximize our odds of identifying facilities, such as consulting a wide variety of diverse sources, but it is still probable that we failed to capture *all* ENR plants that operated globally.

Second, we sometimes found very little information about a plant beyond the fact that it existed. Seemingly straightforward facts, like the exact years that a facility operated, were often difficult to identify. On many occasions, it was also challenging to uncover whether a plant was built with foreign assistance or operated under IAEA safeguards. In coding certain variables, we sometimes had to make educated guesses based on the facts we had at our disposal (we detail those cases in the country coding sheets). Some of our coding decisions may change as new information emerges over time.

Third, there is conflicting information in the literature – particular when it comes to the size of a plant and the years for which it operated. To resolve discrepancies, we made subjective judgments about which sources seemed more credible (we note these cases in the country coding sheets, too). Users of the dataset may therefore notice some instances where the operational years we provide differ from information contained in other sources. In general, we privileged primary documents and databases over secondary accounts. We tended to side with two sources, in particular, when there was conflicting information: the IAEA's INFCIS database and a Pacific Northwest National Laboratory report coauthored by one of our collaborators on the broader project that resulted in the production of this dataset.[4]

Fourth, laboratory-scale facilities present their own unique challenges. We have less information about lab-based activities overall, in part because they are often seen as less significant from a proliferation standpoint. Moreover, as underscored above, it can be difficult to determine when these sites become facilities, according on our criterion. Overall, we are more confident about the completeness and accuracy of information on pilot and commercial plants, compared to lab-scale facilities.

In addition, it is important to underscore that our dataset does not directly measure the amount of weapons grade highly enriched uranium (HEU) or plutonium a country has produced. The operation of an ENR plant – particularly a pilot or commercial facility – is a rough indicator for whether a state can produce fissile material (and therefore a reasonable proxy for nuclear latency). Further, the total number of plants a country has in operation will give one a general sense of an ENR program's scale. However, the operation of ENR plants does not always imply that a state possesses quantities of HEU or plutonium sufficient to make a nuclear weapon. The case of Iraq provides a useful illustration. In 1990, on the eve of the Persian Gulf War, Iraq had several ENR plants operating but it had produced only small quantities of fissile material at that time. In general, though, countries typically have the wherewithal to produce enough fissile material for at least one nuclear weapon once

[4]Zentner, M.D., G.L. Coles, and R.J. Talbert. 2005. *Nuclear Proliferation Technology Trends Analysis.* Richland, WA: Pacific Northwest National Laboratory. Report 14480. Note, however, that neither source contains information on every facility in our dataset.

they have pilot or commercial ENR plants.

Errors and Omissions

The NL dataset may evolve over time, particularly as new information emerges about states' ENR programs. Please send an email to mfuhrmann@tamu.edu if you discover any errors or omissions.