Contract No:

This document was prepared in conjunction with work accomplished under Contract No. 89303321CEM000080 with the U.S. Department of Energy (DOE) Office of Environmental Management (EM).

Disclaimer:

Savannah River National Laboratory®

A U.S. DEPARTMENT OF ENERGY NATIONAL LAB • SAVANNAH RIVER SITE • AIKEN, SC • USA

Hierarchical Conceptual Model Event and Activity Domains for Forecasting State-Sponsored Civil Nuclear Power Activities

T. L. Danielson

E. D. LaBone

March 2022

SRNL-STI-2022-00120, Revision 0

SRNL.DOE.GOV

DISCLAIMER

Printed in the United States of America

Prepared for
U.S. Department of Energy

SRNL-STI-2022-00120
Revision 0

Keywords: *Artificial Intelligence*
Machine Learning
Civil Nuclear Power
Natural Language Processing

Retention: *Varies*

Hierarchical Conceptual Model Event and Activity Domains for Forecasting State-Sponsored Civil Nuclear Power Activities

T. L. Danielson
E. D. LaBone

March 2022

Savannah River National Laboratory is operated by Battelle Savannah River Alliance for the U.S. Department of Energy under Contract No. 89303321CEM000080.

REVIEWS AND APPROVALS

AUTHORS:

____

T. L. Danielson, SRNL, Environmental Sciences & Dosimetry   Date

____

E. D. LaBone, SRNL, Environmental Sciences & Dosimetry   Date

APPROVAL:

____

M. Cofer, Manager   Date
SRNL, Environmental Sciences & Dosimetry

PREFACE OR ACKNOWLEDGEMENTS

The authors acknowledge the Department of Energy's Office of Defense Nuclear Non-Proliferation Research and Development (NA-22) for funding this work.

EXECUTIVE SUMMARY

In FY20, the Savannah River National Laboratory and the Sanghani Center for Artificial Intelligence and Data Analytics (Virginia Tech) entered a collaboration funded by the Department of Energy's Office of Defense Nuclear Nonproliferation Research and Development to develop a machine learning based modeling pipeline to extract proliferation events of interest from open data sources. The prototype modeling pipeline that was developed relies on the use of time dependent word embedding models to identify contextual shifts in key words and phrases that act as indicators of events of interest. The FY20-21 efforts were focused on a narrow topical domain of forecasting "fissile core fabrication" at the Savannah River Site prior to its official announcement in 2018.

In FY22, the research team was funded to continue development of the modeling pipeline by applying it to the problem of forecasting new, and/or significant changes to existing, civil nuclear power reactors around the world. In this effort, the development will focus on proving applicability to a broader topical domain and in data environments that may contain more sparse information, relative to the United States. In this report, the worldwide landscape of civil nuclear reactors is outlined, the timeline of interest is defined, and a hierarchical conceptual model is established to identify the activity domains of interest and the event domains of interest. This preliminary effort will guide the curation of a glossary of key terms for data acquisition, as well as the downstream modeling efforts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABWR	Advanced Boiling Water Reactor
APWR	Advanced Pressurized Water Reactor
BWR	Boiling Water Reactor
FBR	Fast Breeder Reactor
HTGR	High-Temperature Gas-cooled Reactor
PHWR	Pressurized Heavy Water Reactor
PWR	Pressurized Water Reactor
SRNL	Savannah River National Laboratory

1.0 Introduction

In FY20, the Savannah River National Laboratory and the Sanghani Center for Artificial Intelligence and Data Analytics (Virginia Tech) entered a collaboration funded by the Department of Energy's Office of Defense Nuclear Nonproliferation Research and Development to develop a machine learning based modeling pipeline to extract proliferation events of interest from open data sources. In the FY20-21 efforts, the team explored two retrospective text-based data sources consisting of social media (Twitter) and a broad internet archive (Webhose Ltd.) to identify events and indicators that plutonium pit production would be performed at the Savannah River Site prior to its official public announcement in May of 2018. While the team documented success in these efforts [1-3], the prototype modeling pipeline was applied in what is considered an open data environment pertaining to nuclear activities in the United States and to a narrow (i.e., high specificity) nuclear activity domain of fissile core fabrication. In FY22, the team was funded to continue development of the modeling pipeline by applying it to identify events and indicators of new, and/or changes to existing, state-sponsored civil nuclear power reactors in various locations around the world.

Notably, this effort will be applied to a broader topical domain than the FY20-21 modeling pipeline, while still seeking a high degree of specificity for the events that are captured. A successful modeling pipeline will capture events leading to the construction or modification of different *models* of reactors in several different *locations* around the world that are built by *various contractors* and at different *times*. In doing so, the team will seek to demonstrate a generalized applicability of the modeling pipeline which is built on a foundation of natural language processing (more specifically, time dependent word embedding models), graph theory, and anomaly detection to identify contextual shifts in key words and phrases that act as indicators of events of interest.

Prior to acquiring data, the first task to be performed is the development of a hierarchical conceptual model to outline the types of activities and events which might be discovered in the open source. This preliminary modeling step identifies where domain overlap may occur in queried data sources and establishes the nature of events which should be discovered, while also guiding data acquisition. The following sections will provide an overview of the planned, under construction, operational, and shutdown civil nuclear reactors to illustrate the global landscape of civil nuclear power activities at the time of this report. Next, a hierarchical conceptual model defining the activity domains of interest will be presented to provide examples of the breadth of civil nuclear activities that are occurring throughout the world that will be targeted as test cases. Finally, the event domains of interest will be presented to provide an overview of the event types that are anticipated to be discovered.

2.0 Hierarchical Conceptual Model

The demonstration prototype that was developed in FY20-21 will be adapted to a new topical domain of "civil nuclear power" to extract events and indicators of new, and/or changes to existing, state-sponsored civil nuclear power reactors around the world. Notably, the time horizon for the development of a civil nuclear reactor can span many years, where the trajectory of events that occur from inception to start-up to shut down can be diverse and unpredictable when comparing between civil nuclear power reactors built in different locations and for different end-users. A well-informed subject matter expert might define individual stages of the development by key milestone events of interest that vary in nature (e.g., political, financial, technical, etc.). With the completion or discovery of each additional milestone, the possibility to make inferences about an eventual endpoint (or a list of possible of endpoints) becomes more tractable. A successful modeling pipeline will work from a high-level conceptual model to automatically fuse events of interest that are extracted from multiple open data sources as early as possible such that inferences about activities related to civil nuclear power can be made.

The existing prototype modeling pipeline leverages time dependent word embedding models to identify contextual shifts in key words and phrases that are related to the activity of interest. In this approach, a contextual shift in a key word/phrase is indicative of the occurrence of an event of interest (related to the activity of interest) in the preceding time periods. Using language models to extract indicators of any specific activity of interest from the open source faces the challenge of a generally high signal-to-noise ratio that results from the continuously growing and evolving body of information. Therefore, a key step in the modeling pipeline is extracting primarily domain specific data, such that much of the noise from unrelated data is reduced. To accomplish this, a hierarchical conceptual model is outlined that is comprised of "Activity Domains of Interest", where each activity domain is defined by specific "Event Domains of Interest". Having such a conceptual model guides the curation of a limited glossary of key terms/phrases that describe the event domains for each activity and also allows the labeling of key terms to specific events.

In the following sub-sections, the specific timeframe, and locations, to which the modeling pipeline will be applied will be outlined. In addition, a hierarchical conceptual model of "Activity Domains of Interest" and "Event Domains of Interest" will be outlined which will guide data acquisition and downstream modeling efforts.

2.1 Overview Civil Nuclear Reactors Around the World

In FY20-21 efforts, the modeling pipeline was applied to activities in the United States, which is a relatively open data environment. Therefore, a key aspect of the FY22-23 efforts is to apply the modeling pipeline in data environments outside the United States where information may be less publicly available and/or intentionally obscured in different ways. Therefore, the modeling pipeline will be applied to identify events and indicators of activities surrounding civil nuclear power reactors in countries outside the North American continent. Figure 2-1 through Figure 2-4 show the under construction, operational, shutdown, and planned nuclear reactors on continents outside North America across all time[1]. Using these criteria, there are 621 nuclear reactors (capturing multiple reactors at a single site), comprised of 12 different reactor types and 129 different models, in 37 countries.

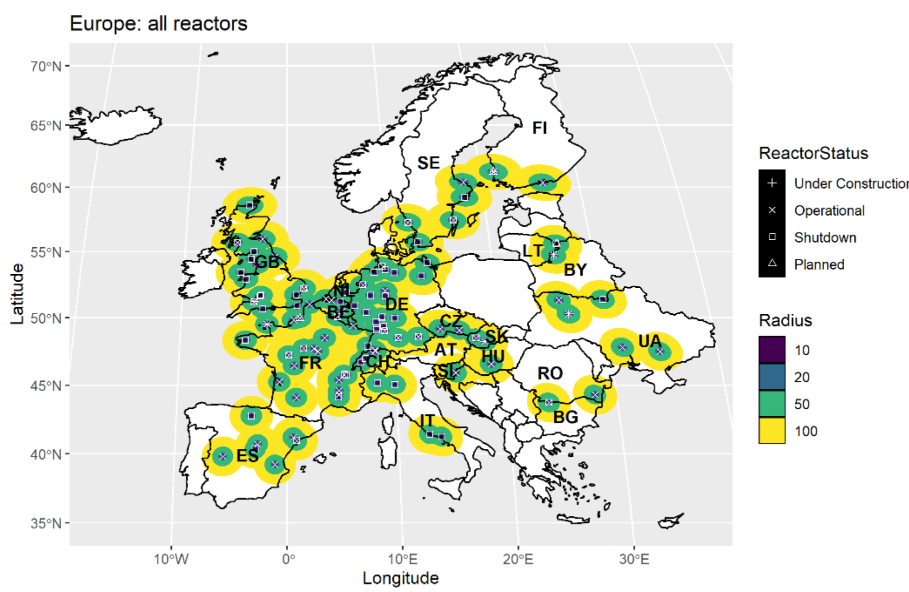

Figure 2-1. Under construction, operational, shutdown, and planned nuclear reactors in Europe.

[1] Worldwide planned, under construction, operational, and shut down nuclear reactor datasets were obtained from: https://github.com/cristianst85/GeoNuclearData

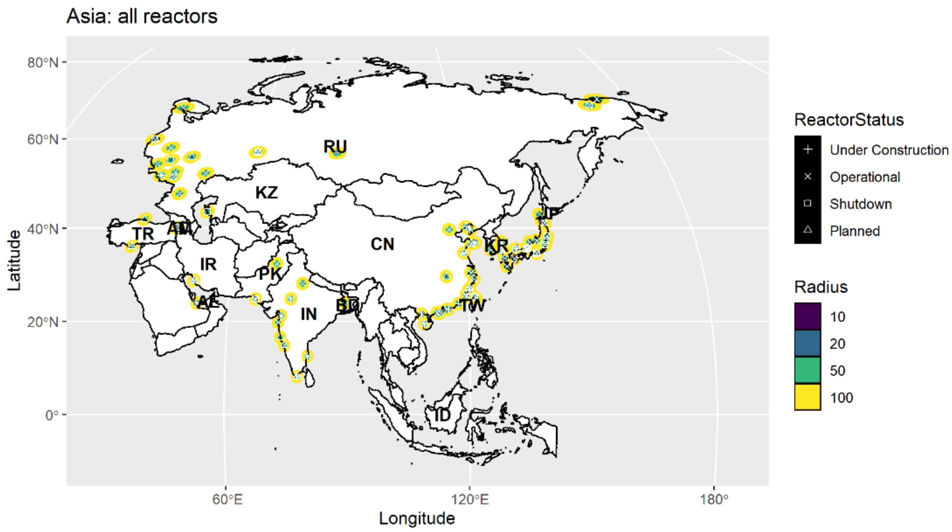

Figure 2-2. Under construction, operational, shutdown, and planned nuclear reactors in Asia.

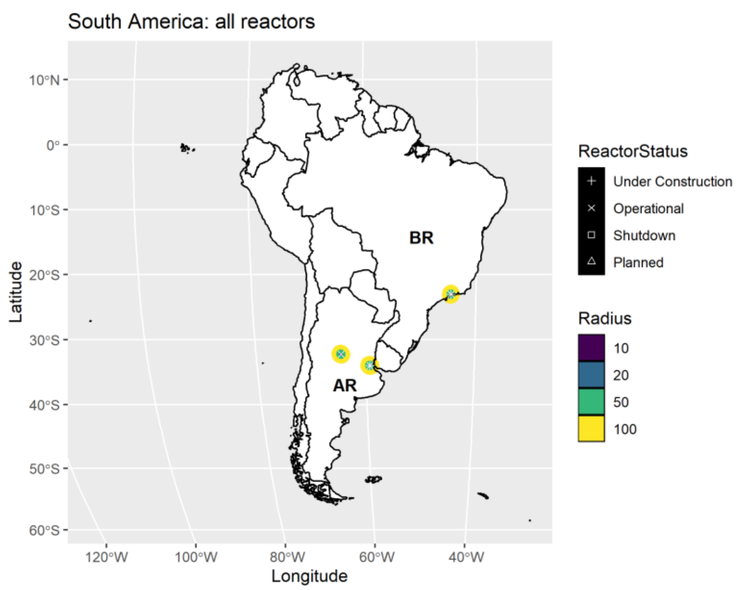

Figure 2-3. Under construction, operational, shutdown, and planned nuclear reactors in South America.

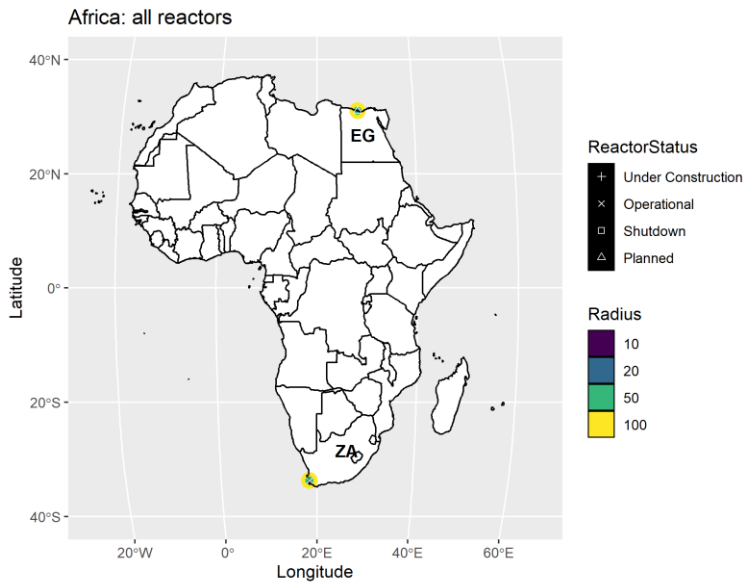

Figure 2-4. Under construction, operational, shutdown, and planned nuclear reactors in Africa.

Data will be gathered from queries placed to a Twitter database and a broad internet archive, as were used in FY20-21. Notably, prior to 2012, the quantity of data available from the retrospective open text-based data sources is limited. Therefore, to identify events and indicators of new, and/or changes to existing, nuclear reactors, the modeling pipeline will be applied only to reactors that are under construction, operational, or planned from 2012 through the present day. Filtering in this manner leaves 205 nuclear reactors in 22 countries with 7 reactor types and 50 different reactor models. The nuclear power reactors that meet this criterion are shown in Figure 2-5 through Figure 2-10. The highest number and density of nuclear reactors are found in Europe and Asia and therefore, the planned reactors on these continents are shown in separate images as well.

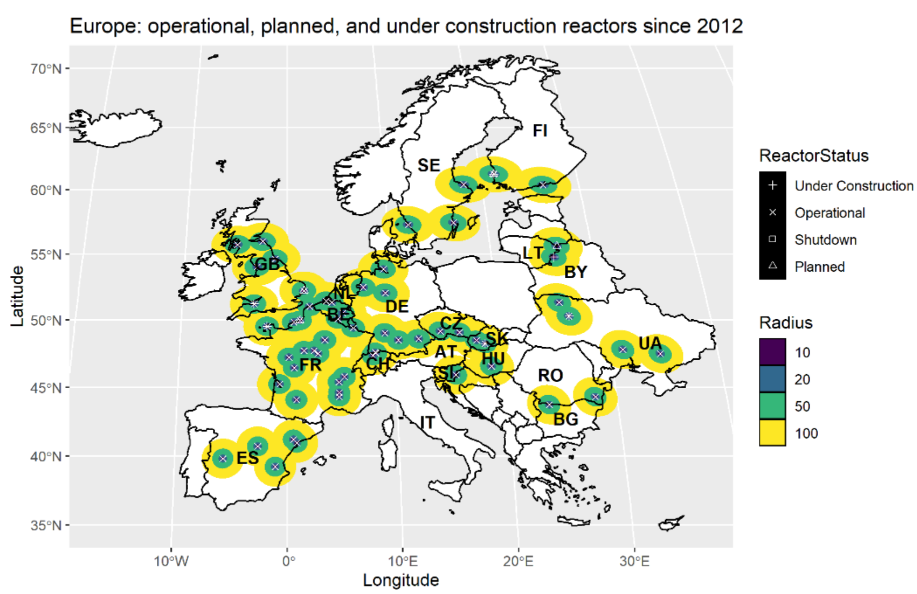

Figure 2-5. Under construction, operational, and planned nuclear reactors in Europe from 2012 to present.

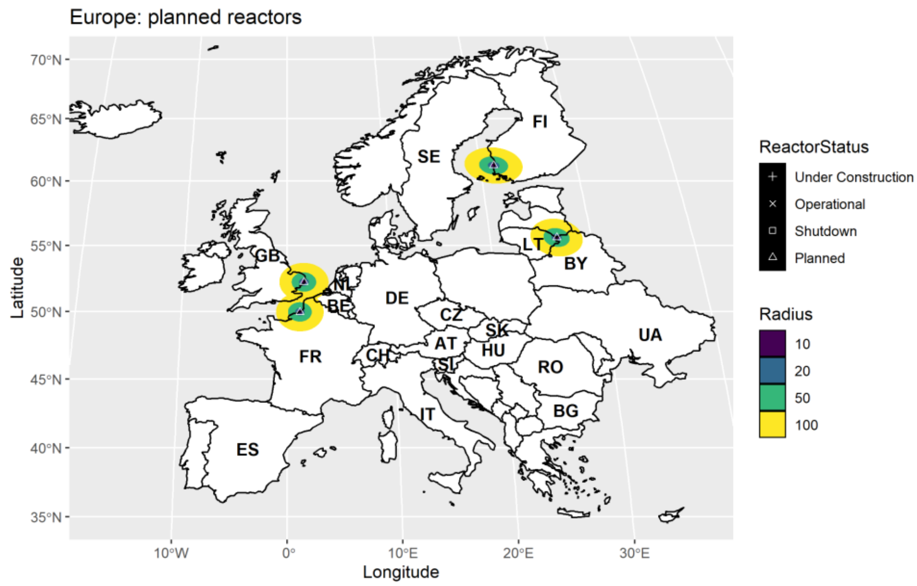

Figure 2-6. Planned nuclear reactors in Europe from 2012 to present.

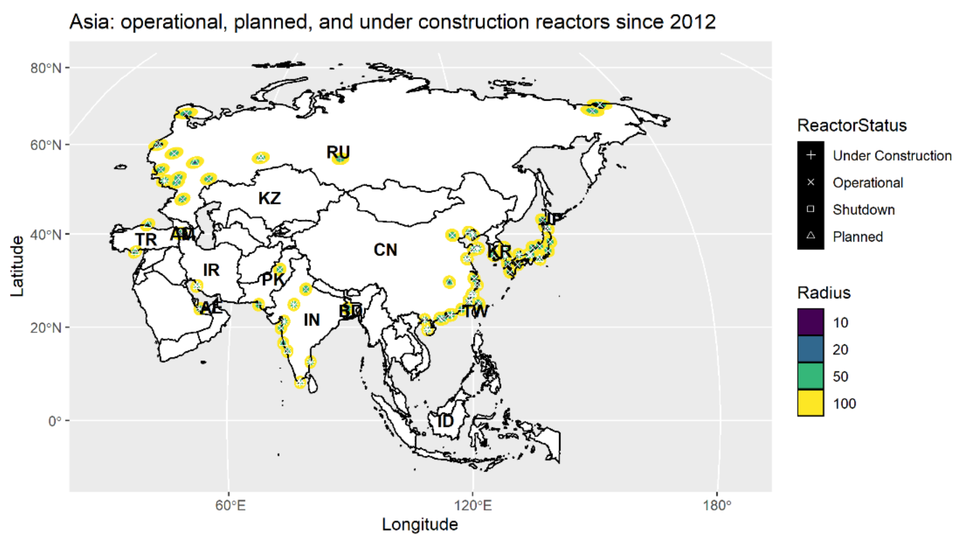

Figure 2-7. Under construction, operational, and planned nuclear reactors in Asia from 2012 to present.

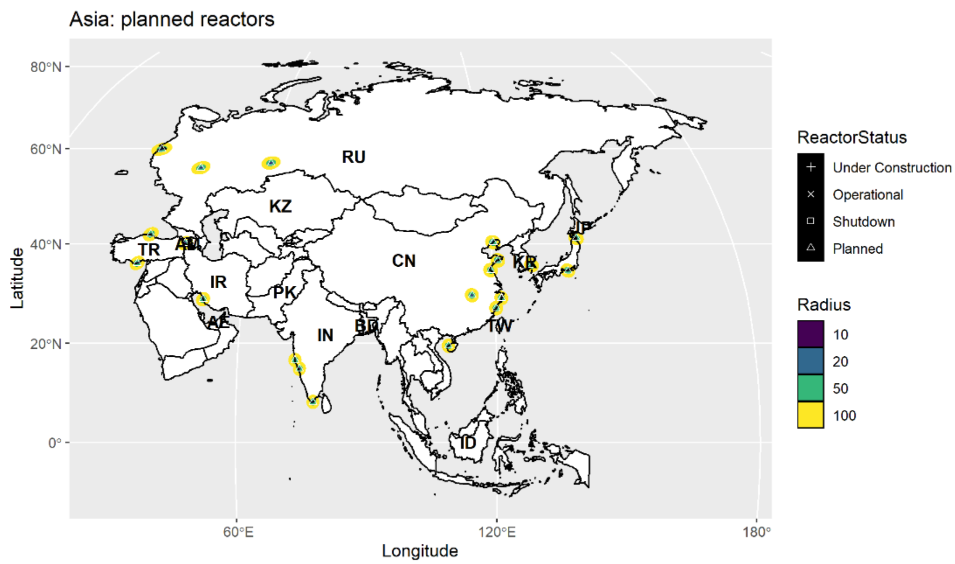

Figure 2-8. Planned nuclear reactors in Asia from 2012 to present.

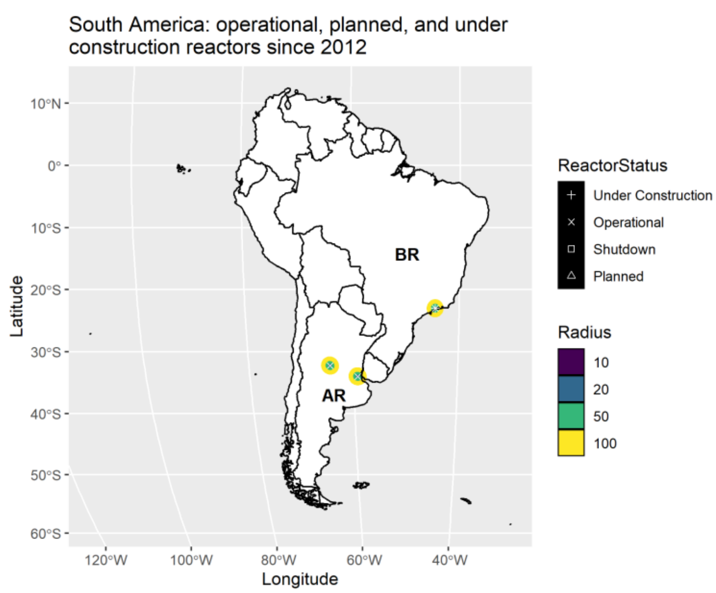

Figure 2-9. Under construction, operational, and planned nuclear reactors in South America from 2012 to present.

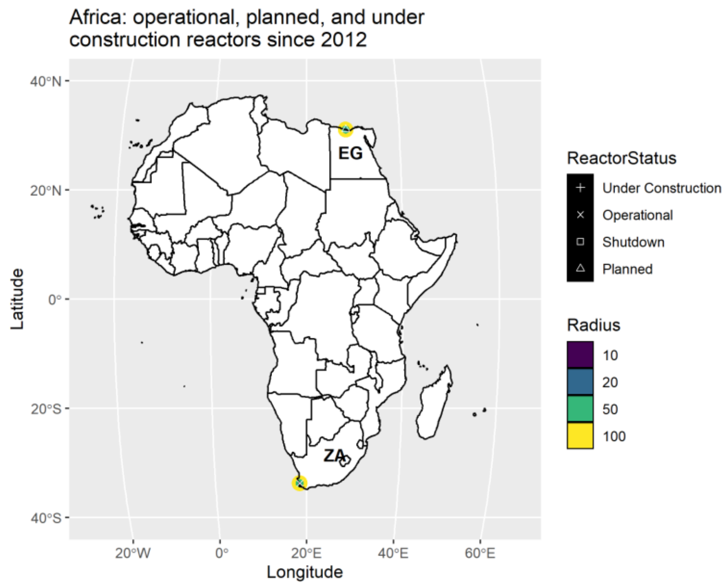

Figure 2-10. Under construction, operational, and planned nuclear reactors in Africa from 2012 to present.

2.2 Activity Domains of Interest

Nuclear activities that occur routinely throughout the world as countries develop nuclear power, nuclear technology/industrial/research, or weapons capabilities have the potential for domain overlap. For example, in the timeline of interest, several different nuclear reactor types (listed in Table 2-1) are represented. Furthermore, each reactor type may have different models that have slight variations in the design and operational parameters tailored to the needs of the end-user. While the design and operational parameters may differ, each reactor type relies on the same basic principle that a sustained fission reaction transfers heat to water to produce steam that drives a turbine to produce electricity. Therefore, there is the potential for significant overlap in the individual components and processes that would be required to construct any of the different types or models of reactors.

Table 2-1. Reactor types represented in the worldwide database in the time period of interest.

Type	Description
ABWR	Advanced Boiling Water Reactor
APWR	Advanced Pressurized Water Reactor
BWR	Boiling Water Reactor
FBR	Fast Breeder Reactor
HTGR	High-Temperature Gas-cooled Reactor
PHWR	Pressurized Heavy Water Reactor
PWR	Pressurized Water Reactor

The steps of the nuclear fuel cycle can differ between reactor types but are broadly illustrated in Figure 2-11. Notably, a subset of the nuclear reactor types that are used for civil nuclear power can also be purposed to produce weapons grade materials. Likewise, there is the potential for overlap in the preliminary steps (i.e., prior to operation) of the nuclear fuel cycle as applied to produce weapons materials and nuclear power materials. Therefore, in general, information regarding the stages leading up to, and including, the operation of the reactor might produce indicators of events of interest that reveal the planning, construction, and operation of a new nuclear reactor. Conversely, information regarding the stages including, and after, the operation (e.g., reprocessing) might reveal indicators of events related to changes to existing nuclear reactors such that the events can be characterized as related to either weapons or civil nuclear power activities.

If a researcher or analyst sought to differentiate between each type of reactor or each stage of the fuel cycle, it might be necessary to take an ontological approach that captures precise aspects of the reactor and the fuel cycle (e.g., components or operational parameters). However, early detection of indicators and events using the existing modeling pipeline does not require such a level of specificity and in fact, could prove too exclusive when curating open-source data. Rather, a high-level glossary of key terms that describe events and activities related to each stage of civil nuclear power and the reactor fuel cycle will be developed to formulate domain-specific queries. In this way, overlapping key terms for the activity and event domains are recognized before any modeling is performed.

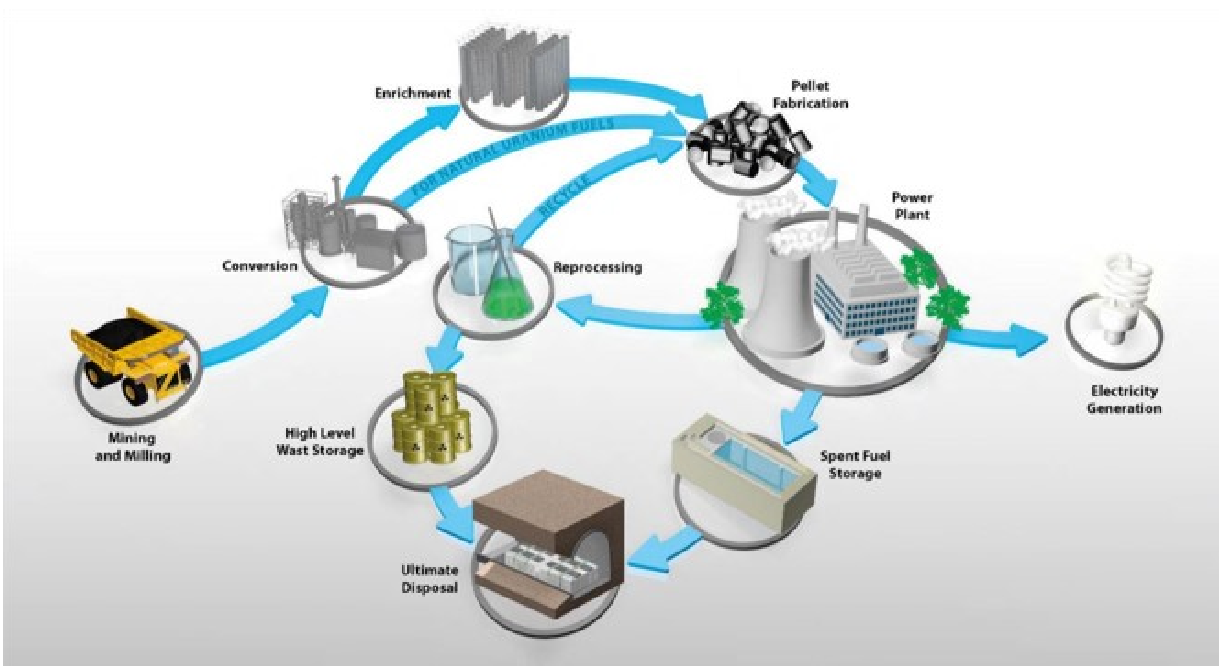

Figure 2-11. Illustration of the steps in the nuclear fuel cycle for civil nuclear power.

The potential for domain overlap is illustrated by the hierarchical conceptual model of nuclear activity domains shown in Figure 2-12 (developed in FY20-21 and extended here for state-sponsored civil nuclear power), where the activities of interest for the current work are shown by solid black outlines. The nuclear activity domains are split into two top-level categories – "Weapons" and "Other" – where a lower vertical positioning indicates higher specificity of the activity domains and activities along the same vertical positioning have greater potential for domain overlap[2]. For example, events related to nuclear materials reprocessing would be expected to be found in both the "Fissile Core Fabrication" and "Civil Nuclear Power" activity sub-domains.

Using this hierarchical conceptual model, the primary focus of this work will be extracting events and indicators related to one of four activity domains: 1) "Planning", 2) "Construction", 3) "Operation", and 4) "Shutdown". Perhaps of most interest are the first three activity domains, which are primarily related to the targeted task of "identifying new, and/or changes to existing, state-sponsored civil nuclear reactors". For example, regarding the planning of a new civil nuclear reactor, the modeling pipeline will seek to

[2] The presence of "Weapons Component Fabrication" and "Fissile Core Fabrication" is to illustrate a comparison in the degree of specificity between the FY20-21 efforts and the FY22-23 efforts, as well as the potential for overlap at that level of specificity.

identify and fuse events that indicate that a specific country or state-level contractor might have interest in building a civil nuclear reactor in a specific location, domestic or international. Regarding the construction of a nuclear reactor, the modeling pipeline might identify events that indicate the timeline of construction of a reactor, or perhaps events that have led to a deviation in some originally identified timeline. Finally, regarding operation, the modeling pipeline might identify events related to operational accidents, suspected operational changes (e.g., re-purposing), or issues in the operation of civil nuclear reactors.

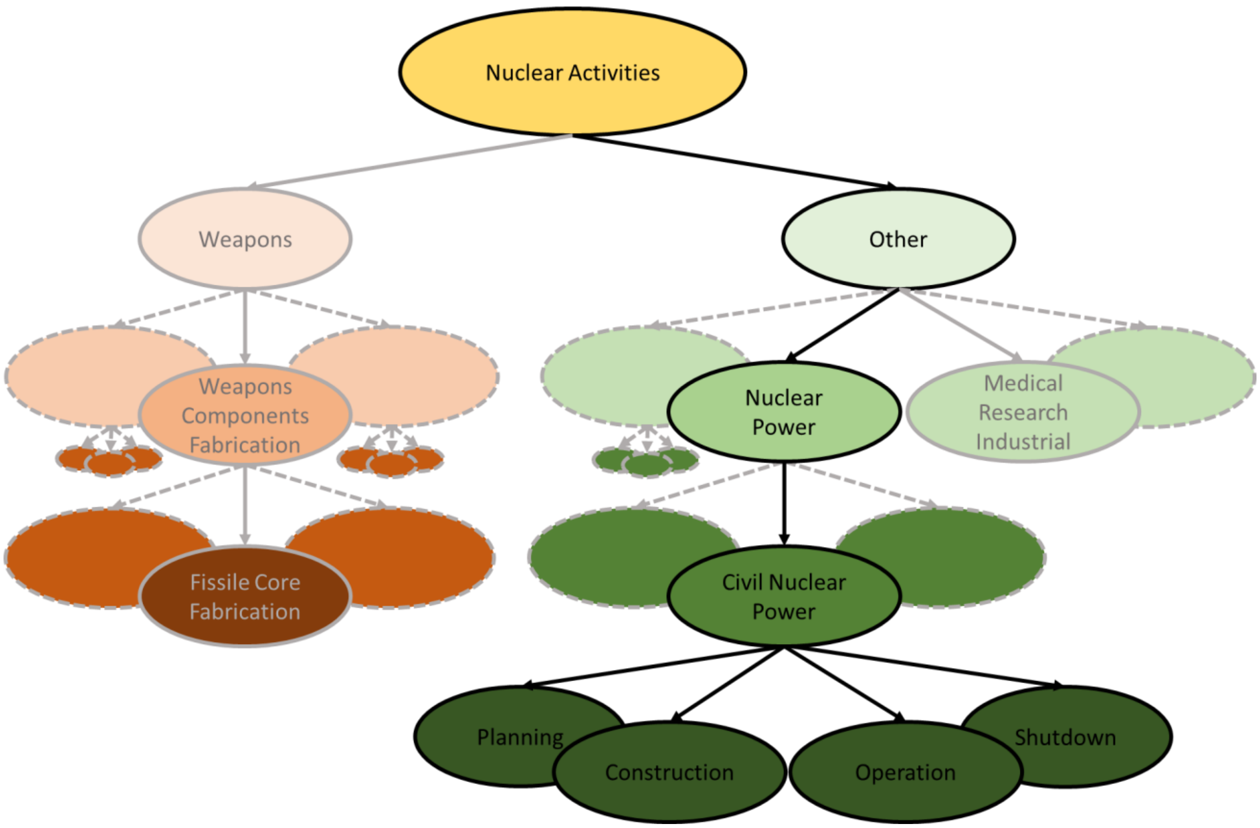

Figure 2-12. Hierarchical conceptual model for the activity domains of interest.

2.3 <u>State-Sponsored Civil Nuclear Power Overview</u>

Many examples of state-sponsored civil nuclear power reactors and activities exist around the world. Perhaps one of the more prolific examples are Russian designed reactors, primarily referred to as VVER reactors (water-water energetic reactor). In many instances (both domestic and internationally), VVER reactors are built, owned, and/or operated by the Russian state-sponsored nuclear corporation, Rosatom. Table 2-2 provides an example list of planned, operational, and under construction VVER nuclear reactors in different locations around the world within the time period of interest.

Notably, a state-sponsored contractor, such as Rosatom, may have several sub-entities that contribute to the life cycle of the nuclear reactor from planning to construction to operation to shut down. For example, Rosatom has different divisions, each existing as a sub-entity for mining, fuel fabrication, engineering, and operation, as well as these activities as executed in different locations. Therefore, references in the open source to any specific activity, event, or entity may reveal connections to an evolving list of relevant entities or sub-entities, where new connections might indicate an event of interest.

The life cycle of each reactor may follow a different trajectory, where varying degrees of geopolitical, technical, and societal events may occur along the way. For example, two countries with a history of diplomacy may be able to plan and construct a reactor faster than countries with no/minimal diplomatic relationship. Furthermore, a history of diplomacy may result in fewer events (political in nature) that could be identified in the open source. Similarly, many of the facilities that have existing reactor infrastructure, but seek to expand with more reactor units at the same site, will likely be identified by a different set of events than an entirely new facility.

Table 2-2. Subset of the planned, operational, or under construction VVER reactors worldwide.

Name	Country Name	Status	Reactor Type	Reactor Model	Construction Start	Operational From
Rooppur-1	Bangladesh	Under Construction	PWR	VVER V-523	11/30/2017	
Rooppur-2	Bangladesh	Under Construction	PWR	VVER V-523	7/14/2018	
Belarusian-1	Belarus	Operational	PWR	VVER V-491	11/8/2013	6/10/2021
Belarusian-2	Belarus	Under Construction	PWR	VVER V-491	4/27/2014	
Tianwan-3	China	Operational	PWR	VVER V-428M	12/27/2012	2/14/2018
Tianwan-4	China	Operational	PWR	VVER V-428M	9/27/2013	12/22/2018
Xudabao-3	China	Under Construction	PWR	VVER-1200/V491	7/28/2021	
Xudabao-4	China	Planned	PWR	VVER-1200		
Tianwan-7	China	Under Construction	PWR	VVER-1200/V491	5/19/2021	
Tianwan-8	China	Planned	PWR	VVER-1200		
El Dabaa-1	Egypt	Planned	PWR	VVER-1200		
Kudankulam-1	India	Operational	PWR	VVER V-412	3/31/2002	12/31/2014
Kudankulam-2	India	Operational	PWR	VVER V-412	7/4/2002	3/31/2017
Kudankulam-3	India	Under Construction	PWR	VVER V-412	6/29/2017	
Kudankulam-4	India	Under Construction	PWR	VVER V-412	10/23/2017	
Bushehr-1	Iran	Operational	PWR	VVER V-446	5/1/1975	9/23/2013
Leningrad 2-1	Russia	Operational	PWR	VVER V-491	10/25/2008	10/29/2018
Leningrad 2-2	Russia	Operational	PWR	VVER V-491	4/15/2010	3/18/2021
Leningrad 2-3	Russia	Planned	PWR	VVER V-491		
Leningrad 2-4	Russia	Planned	PWR	VVER V-491		

Novovoronezh 2-1	Russia	Operational	PWR	VVER V-392M	6/24/2008	2/27/2017
Novovoronezh 2-2	Russia	Operational	PWR	VVER V-392M	7/12/2009	10/31/2019
Rostov-3 (Volgodonsk-3)	Russia	Operational	PWR	VVER V-320	9/15/2009	9/17/2015
Rostov-4 (Volgodonsk-4)	Russia	Operational	PWR	VVER V-320	6/16/2010	9/28/2018
Kursk 2-1	Russia	Under Construction	PWR	VVER V-510K	4/29/2018	
Kursk 2-2	Russia	Under Construction	PWR	VVER V-510K	4/15/2019	
Baltic-1	Russia	Under Construction	PWR	VVER V-491	2/22/2012	
Kalinin-4	Russia	Operational	PWR	VVER V-320	8/1/1986	12/25/2012
Mochovce-3	Slovakia	Under Construction	PWR	VVER V-213	1/27/1987	
Mochovce-4	Slovakia	Under Construction	PWR	VVER V-213	1/27/1987	
Akkuyu-1	Turkey	Under Construction	PWR	VVER V-509	4/3/2018	
Akkuyu-2	Turkey	Under Construction	PWR	VVER V-509	4/8/2020	
Akkuyu-3	Turkey	Under Construction	PWR	VVER V-509	3/10/2021	
Khmelnitski-3	Ukraine	Under Construction	PWR	VVER	3/1/1986	
Khmelnitski-4	Ukraine	Under Construction	PWR	VVER	2/1/1987	

2.4 Event Domains of Interest

The event domains of interest describe the nature of events which might be discovered in the open source that can act as indicators of a specific activity. In FY20-21, a hierarchical conceptual model of event domains of interest was developed which can be applied to all nuclear activities [1]. The event domains of interest are grouped into four primary domains: "Acquisition Events", "Political/Diplomatic Events", "Population Events", and "Economic Events". Each event domain has several sub-domains, as illustrated in Figure 2-13, which ultimately guide the curation of the glossary of terms and allows labeling of the terms according to the top-level event domain. This approach allows events that are extracted from specific documents to be associated with an event type, or multiple event types, having knowledge of which query term returned the document. Likewise, language models can be trained on subsets of documents associated with each specific event domain for potentially higher specificity about any event type.

The discovery of any event type can be in the form of future intent, an ongoing event, or an event that has occurred in the past. Within this framework, the time dependent nature of the modeling pipeline allows the assembly of a timeline of corroborating events from the multiple domains that can be used for inferential assessment. For example, the identification of future intent might allow one to infer planning of civil nuclear power activities. Conversely, the identification of an ongoing event might indicate construction or operation of civil nuclear power activities.

The following sub-sections will describe each event type in more detail and a preliminary assessment of what information may be captured by the event type.

2.4.1 Acquisition Events

The acquisition events domain captures four sub-domains: non-nuclear materials, nuclear materials, nuclear technology, and facility or infrastructural acquisitions. Acquisition events encompass both the potential for future acquisitions, or the discovery that an entity has already acquired the material, technology, or infrastructure. Key terms will be assembled that are associated with acquisition events to identify information that provides the following types of information:

- The intent for new construction
- Ongoing construction activities, known or previously unknown
- The association of construction with specific entities
- Repurposing of existing facilities or the intent to repurpose
- Attempts to obtain nuclear materials (e.g., through mining, purchasing, or reprocessing)
- Attempts or recent acquisitions of parts that would be used for nuclear reactor facilities
- Research and development activities in the field of nuclear power

2.4.2 Political/Diplomatic Events

The political/diplomatic events domain captures three sub-domains: international diplomacy, domestic diplomacy, elections/political appointments. Key terms will be assembled that are associated with acquisition events to identify information that provides the following type of information:

- Changes in diplomatic relations between two countries
- A mutually beneficial driving force for international cooperation with regard to nuclear activities
- A political driving force in favor of nuclear developments within a country
- A political driving force against nuclear developments within a country
- The establishment of new state-sponsored organizations related to nuclear power

2.4.3 Population Events

The population events domain captures three sub-domains: population growth or redistribution, an accumulation of skillsets, and environmental events. Key terms will be assembled that are associated with population events to identify information that provides the following type of information:

- A significant change in a regional population that merits the need for nuclear power
- An increase in the population with expertise in building reactor facilities (e.g., engineers, construction works, etc.)
- Workplace accidents such as significant on the job injuries
- Environmental accidents leading to population hazards

2.4.4 Economic Events

The economic events domain captures three sub-domains: corporate events, economic growth, and infrastructural events. Key terms will be assembled that are associated with economic events to identify information that provides the following type of information:

- The awarding of new contracts
- The involvement of state-sponsored contractors in a specific region
- Significant growth in economic activities that require increased/modified power
- Energy shortages that merit nuclear power

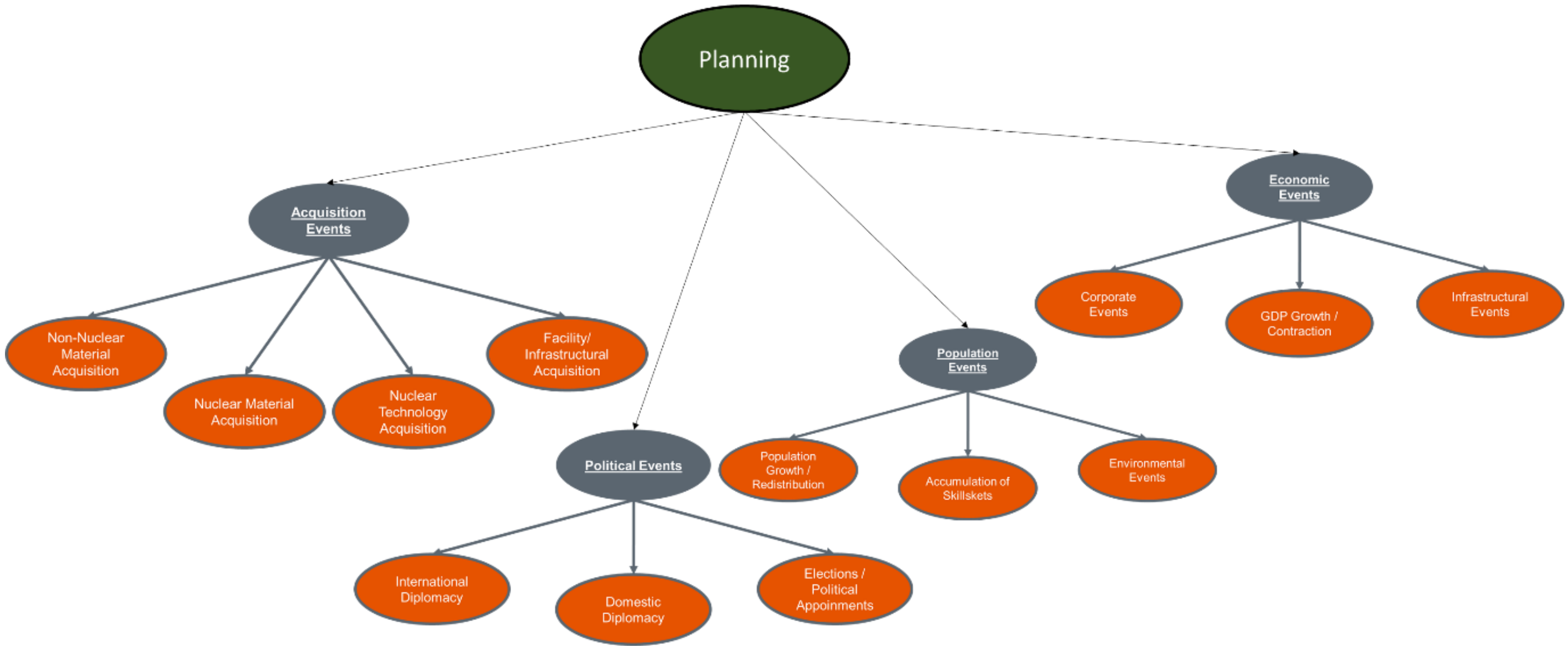

Figure 2-13. Event domains of interest illustrated for the activity domain "Planning".

3.0 Path Forward

The hierarchical conceptual model of activity and event domains will be used to develop a glossary of key terms to obtain time dependent domain-specific data that captures civil nuclear power. The glossary of terms will be developed by identifying open-source documents that capture the nuclear events of interest defined above. While intended to capture domain-specific data, the key terms will be high level to avoid exclusivity of relevant data, given the breadth of the domain of civil nuclear power. For example, the glossary of terms will contain key words and phrases that are related to high-level reactor components (e.g., reactor pressure vessel), construction items necessary for building a reactor (e.g., concrete), common names of reactor fuels, as well as the names of specific entities that are known and closely associated with civil nuclear power activities. The queries will be placed with specific qualifiers to ensure that more general terms are targeted at civil nuclear power.

In the preliminary stages, the modeling pipeline will continue to make use of the Word2Vec-based embedding approach for training time dependent word embedding models, as these models are not context dependent. For example, the word "apple" in the following two sentences – "John Doe is eating an apple." and "John Doe works for Apple." – would have the same word vector. Therefore, with a domain specific dataset, the contextual representation for the word "apple" would undergo significant change if an event occurred (producing many documents or Tweets) where the context of the word shifted from predominantly referring to the fruit versus the computer/software company. This feature of the technique is presumed advantageous in the identification of contextual shifts which act as indicators of events of interest. Context dependent models, such as bi-directional encoder representations from transformers (BERT) models will be explored for use in the downstream modeling efforts for tasks such as question answering.

Following data acquisition, the FY20-21 modeling pipeline will be applied to the datasets to evaluate whether initial queries were successful or additional data is required. This preliminary testing will involve evaluating if the pipeline has identified known activities around the world and that the trained language models have a representative structure (e.g., using principal component analysis of the vectorized corpuses). Additionally, entities that are known to be connected should be recognized by the pipeline.

4.0 Conclusions

A preliminary hierarchical conceptual model has been developed for the FY22-23 efforts, which will apply an existing modeling pipeline to a new topical domain of identifying new, and/or significant changes to existing, state-sponsored civil nuclear power reactors around the world. The hierarchical conceptual model has identified key activity domains and event domains of interest that will be used to curate a glossary of key terms and a domain-specific dataset. The existing pipeline will be applied to two open-source datasets, consisting of a Twitter dataset, containing short-form Tweets, and a broad internet archive, containing long-form news articles. The pipeline will be applied to datasets to identify events of interest in years later than 2012 to the present day and will demonstrate the applicability in data environments outside the United States.

5.0 References

[1] T. L. Danielson, A. A. Kail, and J. A. Pike. April 2020. *Event Definition for the Automated Detection of Nuclear Proliferation Activities.* Aiken, SC: Savannah River National Laboratory, SRNL-STI-2020-00155, Rev. 0.

[2] T. L. Danielson, J. A. Pike, T. Whiteside, B. Mayer, N. Muralidhar, N. Self, P. Butler. January 2021. *Machine Learning Using Open Data Sources for Detection of Nuclear Proliferation Activities (U)*. Aiken, SC. Savannah River National Laboratory, SRNL-STI-2021-00047, Rev. 0.

[3] T. L. Danielson, B. Mayer, N. Muralidhar, J. Miller, H. Dogan, N. Self, P. Butler, F. Liu. January 2022. *Machine Learning Modeling Pipeline for Extracting Nuclear Proliferation Events of Interest from Open Data Sources (U)*. Aiken, SC. Savannah River National Laboratory, SRNL-STI-2022-00036, Rev. 0.

[4] L. G. Worrall. Website: *Safeguards Technology Needs Assessment for Leading Thorium Fuel Cycles*. Oak Ridge National Laboratory. https://www.ornl.gov/division/nsitd/projects/safeguards-technology-needs-assessment-leading-thorium-fuel-cycles.

Distribution:

M. J. Barnes, mark.barnes@srnl.doe.gov
J. S. Bollinger, james02.bollinger@srnl.doe.gov
N. J. Bridges, Nicholas.Bridges@srnl.doe.gov
P. Butler, pabutler@vt.edu
G. R. Cefus, gregory.cefus@srnl.doe.gov
T. L. Danielson, Thomas.Danielson@srnl.doe.gov
C. C. Herman, connie.herman@srnl.doe.gov
C. M. Gregory, clint.gregory@srnl.doe.gov
R. B. James, Ralph.James@srnl.doe.gov
R. D. Jeffcoat, ron.jeffcoat@srnl.doe.gov
B. Mayer, bmayer@cs.vt.edu
N. Muralidhar, nik90@vt.edu
P. R. Nuessle, patterson.nuessle@srnl.doe.gov
N. Self, nwself@vt.edu
D. L. Wilson, david.wilson@srnl.doe.gov
M. Taylor, marc.taylor@srnl.doe.gov
T. P. Taylor, tammy.taylor@srnl.doe.gov
S. J. Branney, sean.branney@srnl.doe.gov
M. Cofer, marion.cofer@srnl.doe.gov
L. T. Brown, lindsay.brown@srnl.doe.gov
B. Lee, brady.lee@srnl.doe.gov
M. Duff, martine.duff@srnl.doe.gov
W. Kuhne, wendy.kuhne@srnl.doe.gov
L. Erikson, luke.erikson@nnsa.doe.gov
C. T. Sloan, craig.sloan@nnsa.doe.gov
Records Administration (EDWS)