ANALYSIS OF 10 CFR PART 810 GENERAL AUTHORIZATION DATA ON

ASSISTANCE TO FOREIGN ATOMIC ENERGY ACTIVITIES

A Thesis

by

LAUREN MARIE POMPILIO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Sunil Chirayath
Committee Members,	Craig Marianno
	James Caverlee
Head of Department,	Yassin A. Hassan

May 2017

Major Subject: Nuclear Engineering

ABSTRACT

The purpose of this research was to implement tools to reliably monitor and analyze incoming Part 810 General Authorization Requests for various nuclear technologies. Texas A&M has utilized and tested the following tools to do so: Excel, Tableau, Apriori and Paterva's Maltego.

Received requests were converted into a searchable format and tagged such that they could be entered into a central repository to enable ad hoc searching. Other exploratory goals are as follows: reporting, analyzing, and predicting when a general authorization request is to be expected. Reporting requirements have been established using the visual analytics software, Tableau, wherein general authorization trends were determined and can be monitored. Prediction has been carried out using Tableau's forecast option, which will anticipate the number of general authorizations to be received by a given country based on prior requests. The Apriori algorithm was also used for prediction purposes. Excel modeling was similarly explored, and it was determined that it should be used sparingly, as unknown variables, such as country and policy needs, are not taken into consideration. Maltego software has been implemented to search the internet and determine when a general authorization report is not received based upon news reports.

Supplementary questions posed by the Department of Energy were explored and answered. A statistical analysis conducted using Tableau, revealed that certain U.S. companies favor specific foreign entities; meaning the U.S. company of interest has

significantly more nuclear technology transfers to one foreign company over others. It was further discovered that China has undergone the General Authorization Process when it is not a specifically authorized destination. Any country not destined as specifically authorized should undergo the specific authorization process; it is not known why this is not already in place. Moreover, certain countries that exceed several general authorizations per year should be monitored to ensure that general authorizations are not being used to circumvent the specific authorization process. The general authorization threshold should be determined by the National Nuclear Security Administration.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sunil Chirayath, and colleague, Dr. Royal Elmore, who brought me in on the project and gave me the utmost support. I would further like to thank Mr. Brent Heilman of Argonne National Laboratory with whom I collaborated on this project and received technical advice and assistance.

CONTRIBUTORS AND FUNDING SOURCES

This work was supported by a thesis committee consisting of Professors Sunil Chirayath and Craig Marianno of the Department of Nuclear Engineering and Professor James Caverlee of the Department of Computer Engineering.

The National Security Science and Policy Institute of Texas A&M University and subsequently myself have received support under the Nonproliferation Analytics Grant from the National Nuclear Security Administration. Los Alamos National Security, LLC manages the NNSA Grant through a Texas A&M subcontract. Primary effort and focus have remained on 810 analytics, specifically in enhancing part 810 categorizations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION AND OBJECTIVE

I.A Background

The Part 810 General Authorization Process is the process set forth by the United States Government in 10 Code of Federal Regulations Export Control (CFR) Part 810. It is important to distinguish between the General Authorization Process and Specific Authorization Process, as countries that undergo the Specific Authorization Process should not be found requesting nuclear technologies, through the General Authorization Process. Unlike General Authorization, Specific Authorization under the authority of section 57.b of the Atomic Energy Act of 1954 requires that for any nuclear material or technology wherein, the persons may engage, directly or indirectly, in the production or development of special nuclear material outside the United States the Secretary of Energy with concurrence of the Departments of State, Defense, and Commerce must approve the request[1]. This provision applies to technology transfers and technical assistance to all activities of the nuclear fuel-cycle, including non-power reactors[1]. Alternatively, General Authorization implements AEA § 57 b.(2), wherein the Secretary grants a general authorization for certain categories of nuclear technologies or materials that are not considered threatening to the interests of the United States. This also includes transfers to specifically authorized destinations given in Appendix A of 10 CFR Part 810[1], which is presented in Appendix I of this thesis. In the case a state, foreign or domestic, desires a nuclear transfer of any sort from the United States, an authorization letter must be sent to the Department of Energy (DOE). The following information is required:

A. Name, address, and citizenship of the applicant, and complete disclosure of all real parties in interest; if the applicant is a corporation or other legal entity, where it is incorporated or organized, the location of its principal office, and the degree of any control or ownership by any foreign person or entity[2].
B. A complete description of the proposed activity, including its approximate monetary value, the name and location of any facility or project involved, the name and address of the person or legal entity for which the activity is to be performed, and a detailed description of any specific project to which the activity relates[2].
C. Any information the applicant may wish to provide concerning the factors listed in 810.10(b)[2].
D. Designation of any information considered proprietary whose public disclosure would cause substantial harm to the competitive position of the applicant[2].

The DOE- National Nuclear Security Administration (NNSA), among other government agencies, accounts for global technological evolutions that are significant challenges to the United States' Export Control System with various export control tools. These tools must simultaneously address nuclear technologies, as well as regulatory requirements. Export controls are increasingly augmented with data mining and statistical analysis to decrease the load on human analysts. A computer based export control program can enhance the ability of the human element by providing insight into the data itself, as well as enable awareness across multiple nonproliferation levels and connections. Computational tools can accommodate various forms of technical knowledge, including insights and social networks, such that no valuable knowledge is missed. Evaluating human export control judgments can provide insight into the historical nature of export control violations, trends, and regulatory weaknesses.

I.B Objective

The objective of this work is to support DOE- NNSA for the following Part 810 General Authorization data objectives:

1. *Searching*: Acquired Part 810 General Authorization data from NNSA will be tagged under agreed upon categories with Argonne National Laboratory. These categories will ultimately be placed in a central repository to facilitate ad hoc searching, wherein new documents will be systematically tagged. Standard searches will be determined and implemented.
2. *Reporting*: Determine what type of reporting is necessary, for example general authorizations to China and, customize as necessary.
3. *Analyzing*: Perform ad hoc analyses of Part 810 General Authorization data to recognize patterns.
4. *Predicting*: Alert and notify 810 officers when a General Authorization report is to be expected and identify using news sources when a General Authorization report is not received. For example, if a news source, such as World Nuclear News, reports that France is building a new research reactor, a General

Authorization request is to be expected and in the case one is not received, a red flag would be drawn.

In addition to the objectives given above, four questions are to be answered regarding the received Part 810 General Authorization Data. They are as follows:

1. Is there any data or given patterns relative to general authorizations for countries that are not designated as Specifically Authorized Destinations? Specifically Authorized Destinations [listed in Appendix I] are those countries that the Secretary of Energy has deemed eligible to undergo the General Authorization Process, as they have non-inimical interests toward the United States.
2. Is there any data or given patterns, relative to sensitive nuclear technologies, such as enrichment or reprocessing [technologies that should not appear in General Authorization Data]?
3. Are there any patterns in technology being sent to a particular country, e.g. same or different technologies and technology changes over time?
4. Are there any patterns exhibited by companies in the United States? Does one country always go to the same one or two U.S. companies or do certain companies only transfer to certain countries or regions?

CHAPTER II

LITERATURE REVIEW

There are traditionally two theories for proliferation and they include "the demand side in which proliferation is driven by the national security demand for the weapons" and "the supply side in which proliferation is driven by the ready supply of nuclear materials and technology."[3] A sole theory, however, is not sufficient to define proliferation and as such there should be interplay between national security motivations and the facilitation that access to supplies gives. Numerous statistical methods have been used in proliferation research and analyses to create a tool to model this interplay.[3] One such method is that of agent based modeling (ABM) and Bayesian Networks. Dr. Elmore showed that using modular Bayesian ABM Nonproliferation Enterprise (BANE) that ABM can successfully model historical cases of proliferation, if agent rules are properly defined[3]. BANE contains an integrated system of AMB and Bayesian methods. "ABM is a computational methodology addressing the uniqueness of those facilitating or preventing the spread of nuclear weapons. Bayesian inference has been employed in fields such as intelligence, where information limits are ever present."[3] Using this methodology data was produced that confirms the assertion that there is a balance between demand/supply drivers and political/military hindrances[3].

Another common proliferation analysis includes increasing the effectiveness and efficiency of the International Atomic Energy Agency's (IAEA) Safeguards System. This includes, but it not limited to, proliferation activities outside a State's declared facilities. Safeguards to measure activities such as these can be quantified with respect to costs and detection probabilities. The problem arises, however, when a facility is undeclared. A risk-driven approach to Acquisition Path Analysis comparing declared and undeclared facilities can be used as a first step analysis for this purpose.[4] The input will include quantification of numerous factors, such as costs of attractiveness values for specific proliferation activities, meaning how enticing an adversary views a pathway to the target of interest. Additional factors include potential safeguards measures and detection probabilities for these measures including the undeclared field.[4] Given that the facility is undeclared, there is a lack of quantification for detection probabilities. To overcome this, a general verification error model can be used alongside an analogy approach. The analogy approach begins by looking at declared facilities, where non-detection probabilities are known. By extension, the same non-detection probability is assumed for undeclared facilities. Say for example it is known that $\beta_{declared} = 10\ \%$, then for the undeclared facility $\beta_{undeclared} = 10\ \%$.[4] This approach, although extremely simple, cannot be validated for the assumed non-detection probabilities.[4]

The IAEA is a major proponent to nuclear non-proliferation. In addition to that mentioned above, they are exploring the potential of open source information in supporting acquisition pathway analysis to design IAEA state level approaches.[5] IAEA safeguards are intended to deter nuclear proliferation and are constantly presented with new challenges. The IAEA envisions that the state level concept be "an objective-based and information driven approach for designing and implementing state level approaches," whose objective is to detect undeclared nuclear material or activities, detect undeclared production of nuclear materials in declared facilities, and detect diversion of declared nuclear material in declared facilities.[5] Under the state level approach, "states will be differentiated based upon objective State-Specific Factors that influence the design, planning, conduct and evaluation of safeguards activities. In designing these state level approaches the IAEA believes that the acquisition path analysis "will identify the plausible routes for acquiring weapons usable material and to assess their safeguards significance."[5] To accomplish this, it must determine plausible acquisition paths, characterize them and finally prioritize. For completion, the acquisition pathway analysis will use open source information. It is important to note that open source is said to be any type of non-classified or proprietary information including media sources, government and non-governmental reports and analyses, commercial data, satellite imagery, and trade data.[5] Table 1 illustrates how open source information could be used in the acquisition pathway analysis. Note, that CSA stands for complementary safeguards agreement and AP represents additional protocol.

Information Collection Areas [4]		Potential Role of Open Source Analysis	Technical/ Official Information Analysis	Media Monitoring	Imagery Analysis	Import/ Export Analysis
Present nuclear fuel cycle	Declared facilities, LOFs, and sites	CSA-only: Corroboration of state declarations (facilities, LOFs) CSA+AP: Corroboration of state declarations (sites)	Y	Y	Y	-
	Exports and imports of nuclear material	Corroboration of state declarations	Y	Y	-	Y
	Nuclear fuel cycle related R&D	CSA-only: main source of information CSA+AP: corroboration of state declarations	Y	Y	-	Y
	Exports and imports of equipment and non-nuclear material	CSA-only: main source of information CSA+AP: corroboration of state declarations	Y	Y	-	Y
	Uranium mines and concentration plants	CSA-only: main source of information CSA+AP: corroboration of state declarations	Y	Y	Y	Y
	Pre-34(c) material holders	CSA-only: main source of information CSA+AP: corroboration of state declarations	Y	Y	-	Y
Past nuclear fuel cycle activities		Corroboration of initial declaration	Y	Y	Y	Y
Planned nuclear fuel cycle activities		Indications of plans to acquire capabilities	Y	Y	Y	Y
Identified anomalies		Indication and investigation of anomalies	Y	Y	Y	Y

Table 1. Roles for Open Sources Analysis for Consolidation in a Given State's Past, Present, and Planned Nuclear Fuel Cycle-Related Capabilities and Infrastructure.[5]

There is a great deal of research that has gone into providing a basis for predicting and evaluating future proliferation events. Per Pacific Northwest National Laboratory, the simplest analysis is to understand and learn from past proliferation events, which includes, "the different paths that have actually been taken to acquire or attempt to acquire special nuclear material."[6] Based on evaluation of historical trends in nuclear technology development it can be predicted the length of time it takes to acquire a technology, the length of time it takes for production of special nuclear material to commence, and the approaches used for acquiring the desired technology.[6]

It is important to understand when and how weapons-usable material can be acquired. This can be carried out using an acquisition network simulation, wherein all the material, facilities, and expertise that are required for obtainment are included and represented by a unique node.[7] The resources required to construct or obtain the choice node can be determined using historical cases and open source information. The most likely pathway that an organization will take can be determined and it considers the resources available to the organization. It considers any of the nodes the organization may already have access to, to determine which path the organization is most likely to find the most attractive.[7]

In addition to acquisition of special nuclear material and predicting proliferation events, the physical protection systems at a facility should be taken into consideration. It is the system in place to prevent and eliminate threats to nuclear materials.[8] Analyses utilizes the adversary sequence diagram to evaluate threat pathways, while also determining the probability that an adversary will be interrupted. In a study conducted at Texas A&M University, a hypothetical 5 MW pool-type reactor that utilizes highly enriched uranium was assumed.[8] The protection layers were as follows: offsite, protected area, controlled building, reactor containment, and reactor equipment room. The most important aspect is the paths of attack for the insider threat.[8] To determine this, the detection and delay components of the physical protection system need to be measured and are subsequently used to construct the adversary sequence diagram.[8] The probability that the adversary is interrupted is then determined by the detection probability of each detection apparatus along the most vulnerable path to the critical detection point.[8] Along this path there is a constant probability of neutralization. These methods were applied to a sabotage scenario at the aforementioned reactor and yielded results that demonstrated the applied methodology to evaluate the vulnerability of a physical protection system are successful for an insider-outsider collusion threat. The probability of interruption decreased significantly for insider-outsider collusion cases, which leads to a higher security risk.[8]

When political or economic measures prove unsuccessful in preventing a country from acquiring nuclear weapons, military action could become necessary. Research performed at the Naval Postgraduate School, has developed a tool to support strategic delay of development of nuclear weapons.[9] The optimal interdiction model selects points for interdicting the weapons program of the country in question. Selection points are chosen such that the induced delay in completing the project is maximized.[9] When run, "results obtained show that the optimal interdiction activities can be solved with an operations research approach. In a few of the multiple activity interdiction runs the activities selected for interdiction were not on the original critical path, and thus an interdiction induces both a delay and a new critical path. These results show the critical path may not be the best source of activities to interdict."[9]

Based on the literature review, even though multiple ways of analyzing proliferation data exists, a methodology to systematically analyze Part 810 General Authorization has never been attempted, which is the topic of the current study.

CHAPTER III

METHODOLOGY

III.A Digitization Process

To meet objectives, Part 810 General Authorization Data was collected over approximately 25 years and includes 12,931 authorization letters. Optical Character Reader (OCR) software transforms General Authorization Data received into a workable form, such that various searches can be performed. There are numerous open source OCR software available. Three OCR software packages were considered for this project. They are as follows: Google Drive OCR, OCRupus, and Tesseract. Google Drive OCR works via upload command, however, there are various rules of thumb. Firstly, the file should be high resolution with clear contrasts and even lighting. Secondly for optimal recognition the text being scanned should be horizontal and read from left to right with standard typefaces, such as Helvetica and Times New Roman[10]. OCRupus is a collection of document analysis programs rather than a turnkey OCR system[11]. At the onset, OCRupus used Tesseract as an internal recognition engine, but later developers redefined the software to run under their own engine, which was described as innovative, but lacking in maturity[12]. Tesseract is a command line tool and produces arguably the most accurate results[13]. Despite the rules of thumb mentioned above, such as high resolution with clear contrasts and even lighting, Google Drive OCR was chosen, as several PDF's were successfully transformed into a workable form, a form in which various in-text searches can be carried out.

III.B Data Analytics

Numerous tools were considered for data analysis at the macro level. Macro level assessment was chosen over micro level, as micro level would only focus on few parts. For example, it would only look at suppliers and technology, while macro level incorporates the larger nuclear proliferation picture focusing on all the following[14]:

A. Suppliers: countries or companies providing technology.
B. Technology: specific items or commodities, or groupings of items.
C. Consumers: countries and/or companies interested in acquiring technology.
D. Economy: the social system focused on production, distribution, and use of technology, both dual use and proliferation only. The level of granularity or detail can vary from company to country, region, or worldwide based on the available data.
E. Proliferation context: organizational framework describing the entire proliferation cycle.

The first analytical tool considered was a Bayes Net, which portrays a graphical representation of relationships among several variables. Mathematically speaking, it is a representation of the joint distributions among random variables, with directed and acyclic links. For example, Bayes Nets are commonly used to support proliferation assessments in determining the likelihood that a state will pursue a nuclear weapon. The model accounts for various social factors, such as political, economic, nuclear capability, security, and national identity. Pacific Northwest National Laboratory (PNNL) has developed a network such as the one described above, wherein information is used from both traditional safeguards and the strengthened safeguards associated with the Additional Protocol to indicate countries with a high risk of proliferating nuclear weapons[15].

A Distance Model was the second tool explored for use. A model such as this captures the net effect of all factors that make certain 'transactions' less likely in each setting. These settings represent commercial relationships or potential relations of the players in the transactions[14]. To review the massive amount of data received, a big data analytics tool, developed by PNNL was reviewed[16]. T-Rex is a visual analytics tool that allows analysts to explore tabular data sources, to quickly identify patterns. The most useful facet of this tool, as it applies to the project at hand, is the timeline view that shows temporal patterns in the data[16].

System Dynamics Models (SD) are extremely useful in understanding the impacts of policy and individual decisions on the behavior of a system. They represent cause and effect relationships; ultimately aiding in the understanding of how the decisions of multiple entities interact to produce an overall behavior that would be nearly impossible to predict in the absence of the model[14]. Coupling a SD methodology with another modeling approach covers its limitations in areas such as treating coarse data in a more granular and precise fashion.

Agent-Based Modeling (ABM) encompasses autonomous decision-making agents that assess the situation, make decisions and execute the appropriate behavior[14]. Justifications for choosing a coupled SD model are presented below:

I. Advances the understanding of the present export control licensing realm including process efficacy.
II. Helps in the discovery of better licensing protocols through simulations of various licensing scenarios.
III. Improves other methods and tools for nuclear proliferation analysis.
IV. Enhances the understanding of nuclear proliferation through various scenarios.

The particular ABM and simulation explored, but not used is entitled Repast HPC. HPC stands for high performance computing. Repast HPC was developed at Argonne National Laboratory. The Repast Symphony toolkit's principles and concepts went into Repast HPC for larger scale simulations. Repast HPC is written in C++ using the message passing interface (MPI) for parallel operations, i.e. wherein numerous processes are running in parallel and memory is not shared across these processes, while also making use of the boost library[17]. In this case, the agents are implemented as objects in C++ classes, while their states are represented by the field variables of the specific classes in question and agent behavior by methods in those classes. Repast HPC is intended to smooth the path from small-scale simulations to large-scale distributed simulations using a Logo-like system[18].

The agent population in Repast HPC is enveloped in contexts that have projections associated with them. A Projection imposes a relational structure on the agents in the Context. Repast implements 3 types of projections: A grid, a continuous space and a network. For example, a grid projection puts agents into a grid matrix structure where each agent is placed in a cell location on the grid. Any agents added to the Context become accompanying vertices to the entire network projection[17]. Thus, a SD model coupled with other methods can prove to be a powerful tool, however, the Part 810 General Authorization data used as part of this research will not be considered under the impact of policy. Thus, ABM was solely used.

This research was done in partnership with ANL, as requested by NNSA. Upon recommendation, the software tool known as Tableau was used as the primary tool for exploring questions posed in the Objective. Tableau is a software commonly used for big data analytics in addition to being highly sought after for advanced sorting capabilities. It combines a structured query language with a graphic interface descriptive language called Visual Query Language (VizQL)[19]. Tableau will query relational databases, cubes, cloud databases, and spreadsheets[19]. Trend lines can be determined between report date and requested patterns of interest if applicable or warranted.

III.C Data Mining

Three tools were considered for data mining purposes. Lumify was the first to be examined. The program is a web-based open source interface[20]; it allows the user to explore relationships in the user's data via a suite of analytic options, including 2D and 3D graph visualizations, full-text faceted search, dynamic histograms, interactive geographic maps, and collaborative workspaces shared in real-time[21]. Lumify is based on five key concepts and they include the following[21]:

1. Ontology- Structure for organizing information to be analyzed.
2. Entity- What is to be represented.
3. Relationship- Link between two entities.
4. Properties- Data encompassing an entity.
5. Graph- Relationship between entities.

Orange Data Mining is a component-based software suite that is free of charge. It enables one to perform simple data analysis, explore statistical distributions, box plots and scatter plots or delve further using decision trees, hierarchical clustering, heat maps, MDS and linear projections[22]. Data mining is carried out through Python scripting with a hierarchically organized toolbox of data mining components and visual programming[23]. The component hierarchies are as follows[23]:

1. Data management and preprocessing
2. Classification
3. Regression
4. Association
5. Ensembles
6. Clustering
7. Evaluation
8. Projections

Ultimately, Maltego was determined to be the most viable option, as the program is most rigorous in internet searches. There are three forms of Maltego and they are as follows: Classic, XL, and CE. Maltego Classic is the chosen form, as it can identify nearly 10,000 sources. The data mining tool produces graphs for link analysis, as well as act as an effective tool for finding relationships between information from various sources on the Internet. This is carried out under the idea of transforms that automate the process of querying different data sources, ultimately displaying results on a node based graph for link analysis. The specific transform to be used is the phrase transform, wherein a key phrase or part thereof is entered and searched for on various websites. Connections are found using open source intelligence techniques by querying sources DNS records, whois records, search engines, social networks, various online APIs and extracting meta data[24].

III.D Weka Data Mining and the Apriori Algorithm

The agent based modeling software entitled Weka, was used to determine association rules for data mining. The program, developed at The University of Waikato in New Zealand, is a collection of machine learning algorithms, including Bayes Net and Apriori. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes[25].

The Apriori Algorithm or Association Rule Mining is defined as being a tool that will, given a certain number of transactions, find rules or prescribed guides that will predict the occurrence of an item based on the occurrences of other items in the itemset[26]. The algorithm is parallelized and implementable. Important definitions are as follows[26]:

1. Itemset: Collection of one or more items; Example: {Recipient Country, Specific Transfer, Recipient Company}.
2. Support count (σ): Frequency of occurrence of an itemset.
3. Frequent Itemset: An itemset whose support is greater than or equal to a *minsup* (minimum support) threshold, as denoted by L_k for the k^{th} itemset.
4. Support (s): Fraction of transactions that contain both X and Y.
5. Confidence (c): How often items in Y appear in transactions that contain X

Key concepts of the Apriori principle are given below[27]:

1. Any subset of the frequent itemset must be frequent.
2. A joint operation is the process of determining the minimum support threshold, wherein a set of candidate i-itemsets is generated by uniting L_{k-1} with itself.

The association rule itself is an implication expression in the form of $X \rightarrow Y$, wherein X and Y are itemsets. This expression is similar to that of a chemical equation, while it does not indicate X approaching Y. For example, the following expression and itemsets are given: $\{Recipient\ Country, Report\ Year\} \rightarrow \{Specific\ Transfer\}$.

Number	Items
1	Recipient Country, Report Year, Specific Transfer
2	Recipient Country, Report Year
3	Report Year, Recipient Country
4	Specific Transfer

Table 2. Items for Example Support and Confidence Calculations.

Support and confidence evaluations are illustrated below for $\{Recipient\ Country, Report\ Year\} \rightarrow \{Specific\ Transfer\}$.

$$s = \frac{\sigma\{Recipient\ Country, Report\ Year, Specific\ Transfer\}}{Number} = \frac{1}{4} = 0.25$$

$$c = \frac{\sigma\{Recipient\ Country, Report\ Year, Specific\ Transfer\}}{\sigma\{Recipient\ Country, Report\ Year\}} = \frac{1}{2} = 0.5$$

The Apriori Algorithm within Weka will determine association rules as illustrated below.[27] Refer to definitions, presented on page 17.

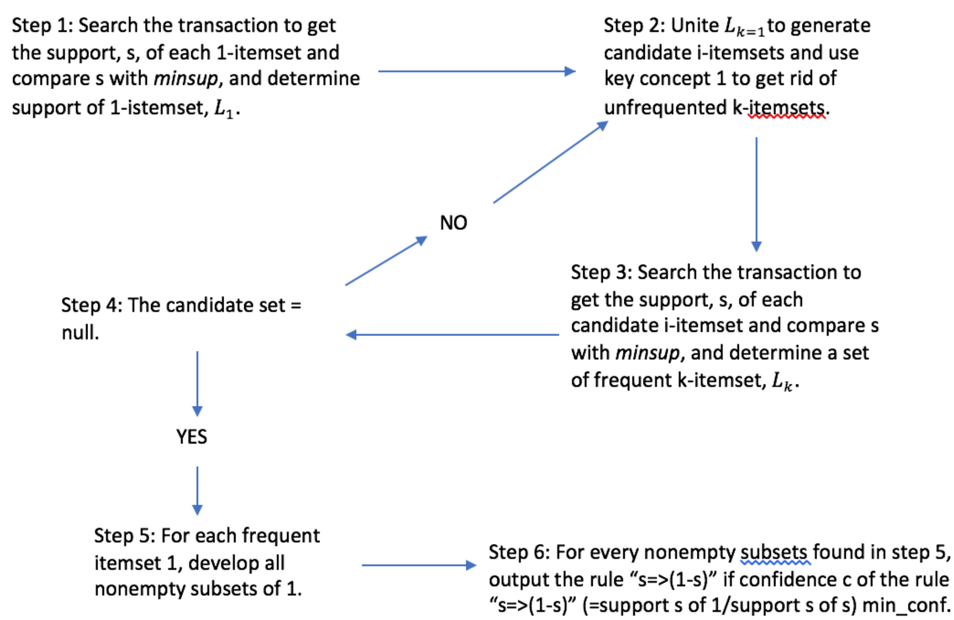


Figure 1. Algorithm Steps to Determine Association Rules for Data Mining.

CHAPTER IV

RESULTS AND DISCUSSION

IV.A Excel Spreadsheet Categorization

Information obtained from the OCR converted General Authorization Data was separated into unified categories in an excel spreadsheet. This database has 12,931 authorization requests. Proper categorical names were chosen that are universally defined and understood. The categorical names can be consistently applied by various 810 workers who employ and use the Part 810 General Authorization Data. The chosen categories are as follows:

1. File Number
2. File Name
3. Report Type
4. Report Date
5. Company Submitting the Report
6. Name of the Individual Submitting the Report
7. Recipient Company
8. Name of the Individual at the Recipient Company (if applicable)
9. Country of Recipient
10. Type of Transfer E.g. Engineering Services, Engineering Support, Technical Information Sharing, and Personnel
11. Specific Transfer E.g. Medical Isotopes Production System, Eddy Current Testing, etc.

12. Begin/End Transfer Dates
13. Foreign National Name, Citizenship and Employment (if applicable)
14. 10 CFR 810.2

It is important to note that 10 CFR 810.2 methodology was used to tag documents based on Specific Transfer and activity description. Possible tagging includes:

1. Uranium conversion (Conversion of U_3O_8 to UF_6)
2. Thorium conversion (Conversion of Th_{232} to $U_{233,}$ addition of F_9 gas converts UF_4 to UF_6)
3. Plutonium conversion (Conversion to weapons-grade plutonium, formation of PuO_2 for MOX fuel)
4. Neptunium conversion (Formation of NpO_2 for MOX fuel)
5. Fuel Fabrication
6. Isotopic Separation
7. Reactors
8. Production ADS
9. Heavy Water
10. Reprocessing

An excel spreadsheet with the above information acted as the central repository to facilitate ad hoc searching, wherein new documents were systematically tagged and standard searches were determined and implemented.

IV.B Answered Questions Presented to DOE

As given earlier, the following four questions were answered:

1. Is there any data or given patterns relative to general authorizations for countries that are not designated as specifically authorized destinations?
2. Is there any data or given patterns relative to sensitive nuclear technologies, such as enrichment or reprocessing?
3. Are there any patterns in technology being sent to a country e.g. same or different technologies and technology changes over time?
4. Are there any patterns exhibited by companies in the United States? Does one country always go to the same one or two U.S. companies or do certain companies only transfer to certain countries or regions?

Inquiry one is analyzed as below:

Non-specifically authorized destinations were found to have general authorization requests and are presented in Figure 2 alongside the number of accompanying requests. Specifically Authorized Destinations are given in Table 9, Appendix I, as they are the participants in the Nuclear Suppliers Group. Non-specifically authorized destinations should be undergoing the specific authorization process; it is not known whether this is checked by the NNSA.

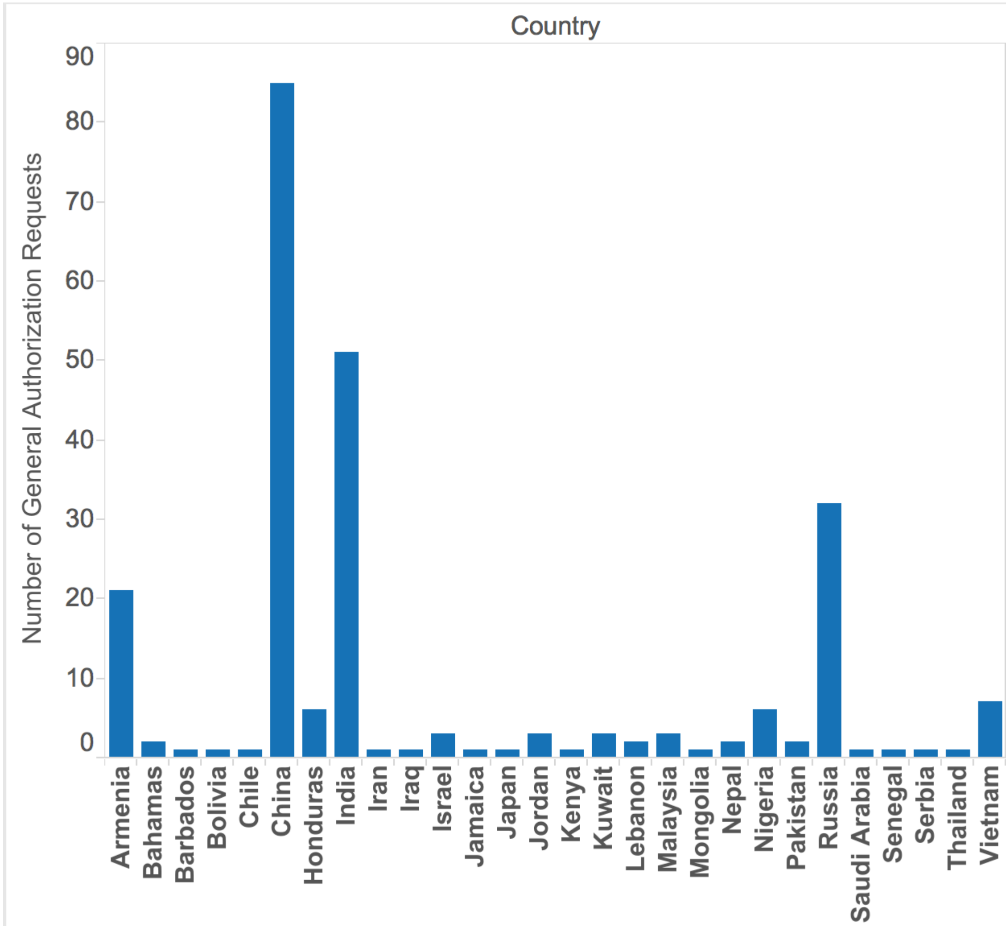


Figure 2. Non-Specifically Authorized Destination Requests and Number of Records.

To determine specific transfer trends to these destinations shown in Figure 2, the number of records over all countries were counted and summed for employment, engineering services, engineering support, personnel use, and technical information sharing (Figure 3). Engineering services would include reactor part replacement, engineering support encompasses funding, personnel use includes; but is not limited to; internships and technical information sharing includes reactor part specifications.

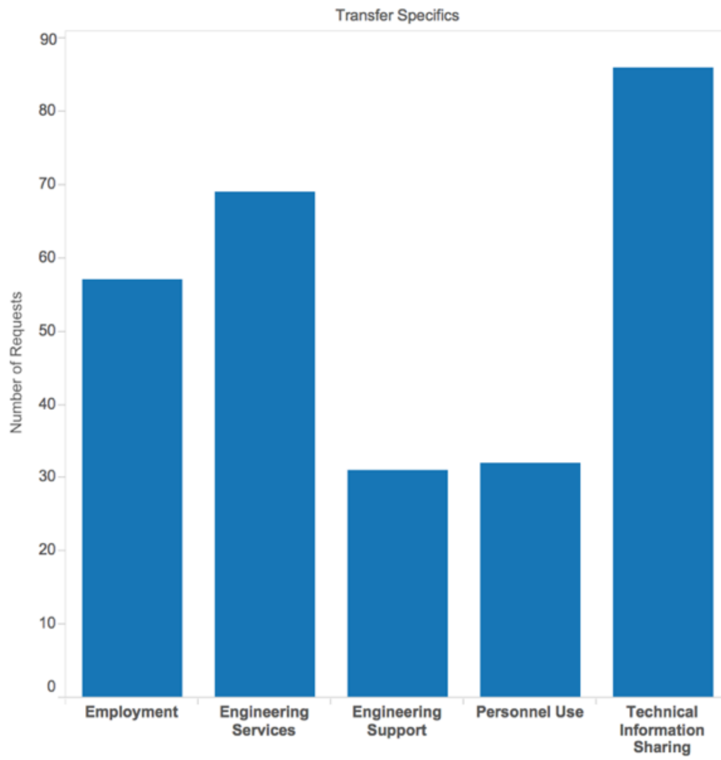


Figure 3. Specific Transfers Summed Over All Non-Specifically Authorized Countries.

Most records revolve around technical information sharing. It is important to note, however, the spread between the maximum and minimum points. The maximum and minimum points for the Number of Requests are approximately 85 and 30; a significant difference, indicating that engineering support and personal use are a minor desire of non-specifically authorized destinations. Technical information sharing, employment, and engineering services, on the contrary, should be closely monitored. Moreover, non-specifically authorized countries that exceed a certain threshold of general authorizations should be supervised to ensure that these countries of interest are not using multiple general authorizations to circumvent specific authorization. To determine trends, the number of authorizations as pertaining to the above specific transfers were counted and plotted for non-specially authorized countries. Figure 4 illustrates the trends for engineering services, while Figures 5, 6, and 7 show engineering support, personnel and technical information sharing. All figures were created using Tableau and as such there are preset settings that place axis titles in specific locations.

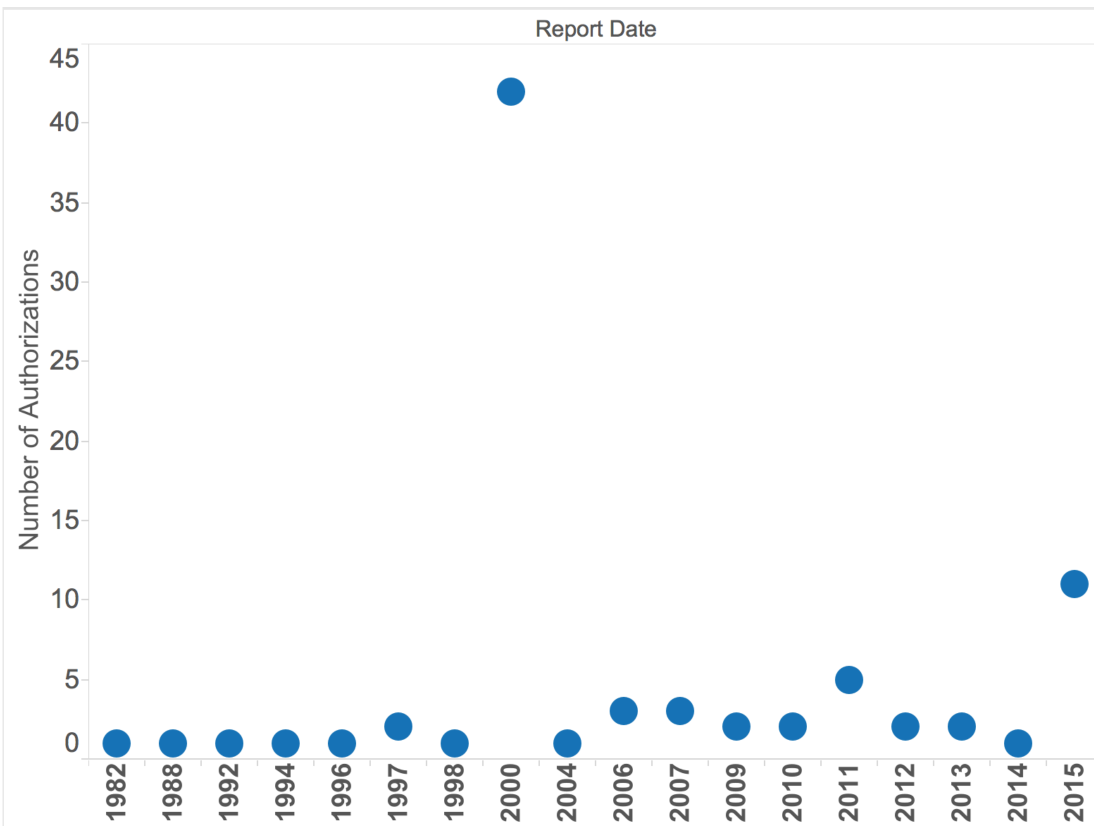


Figure 4. Number of Authorizations for Engineering Services as a Function of Time.

There isn't a specific trend to be identified. The number of authorizations plateaus and proceeds to peak in 2000, then decreases in the following years. The number of authorizations is dominated by China and Armenia in 2000. Both countries have 20 General Authorization Requests. Armenia solely requested information related to operational safety, while China only requested specifications for natural gadolinium oxide powder.

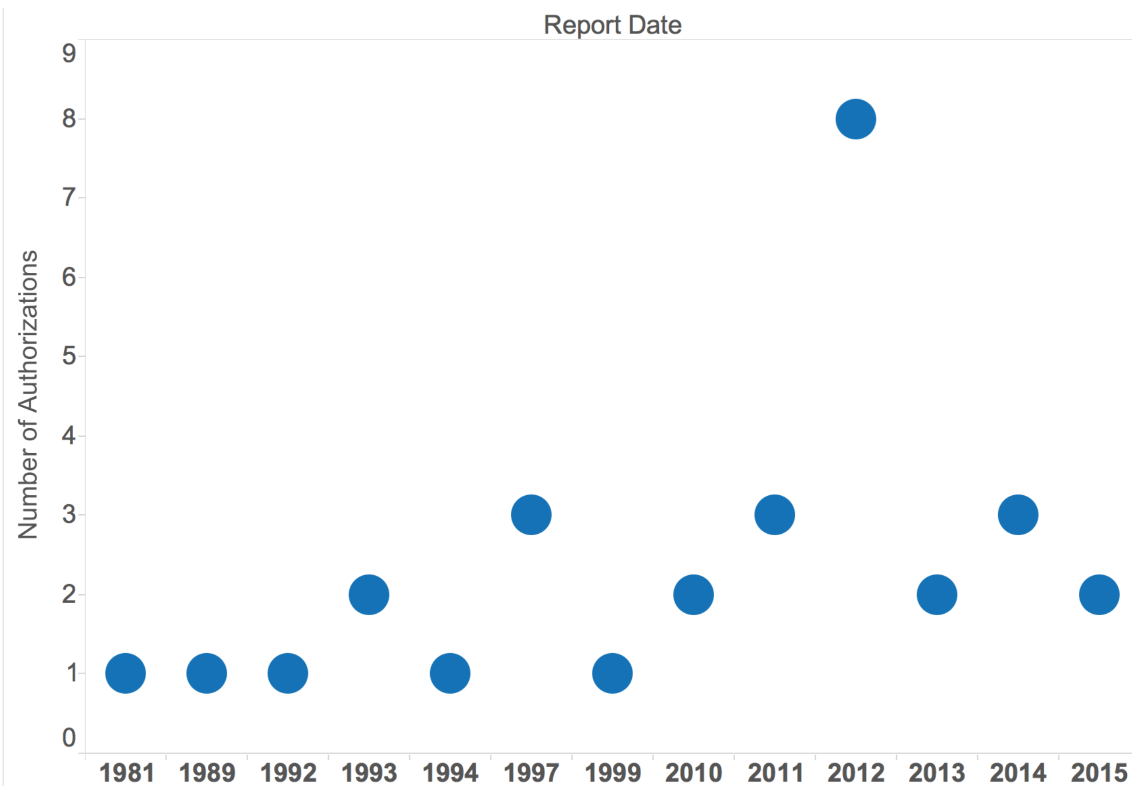


Figure 5. Number of Authorizations for Engineering Support as a Function of Time.

As illustrated, the number of authorizations pertaining to engineering support fluctuate over time and peak in 2012. As with engineering services there is no trend to be identified. China and India dominate the number of authorization requests; China has 5, while India has 2.

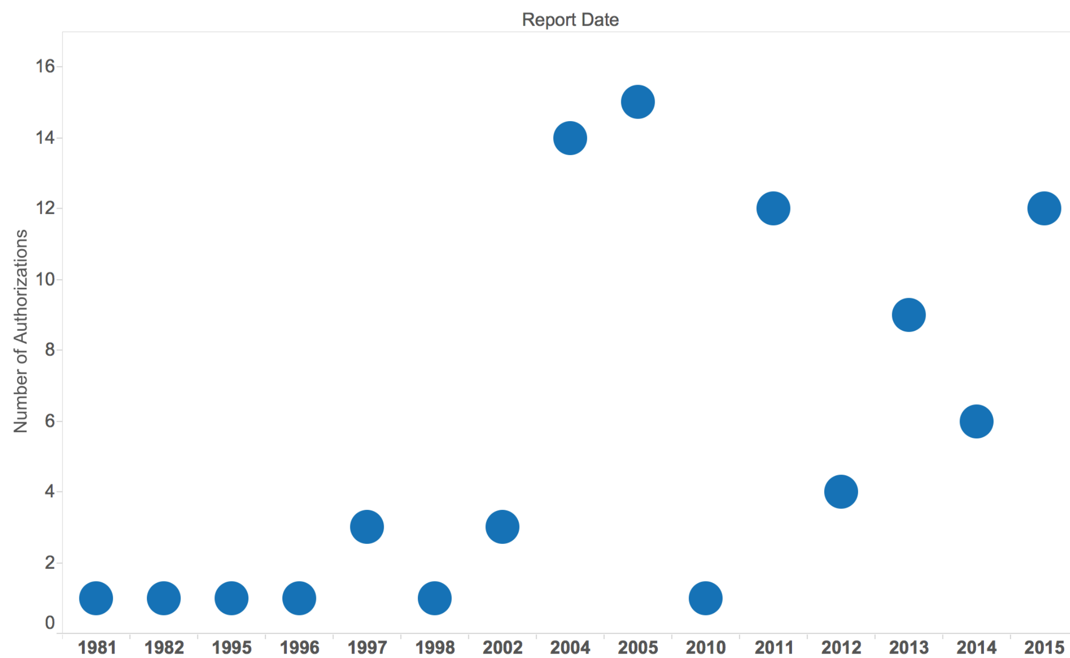


Figure 6. Number of Authorizations for Personnel as a Function of Time.

As shown in Figure 6, the number of authorizations for personnel fluctuate for non-specifically authorized destinations, peaking in 2005.

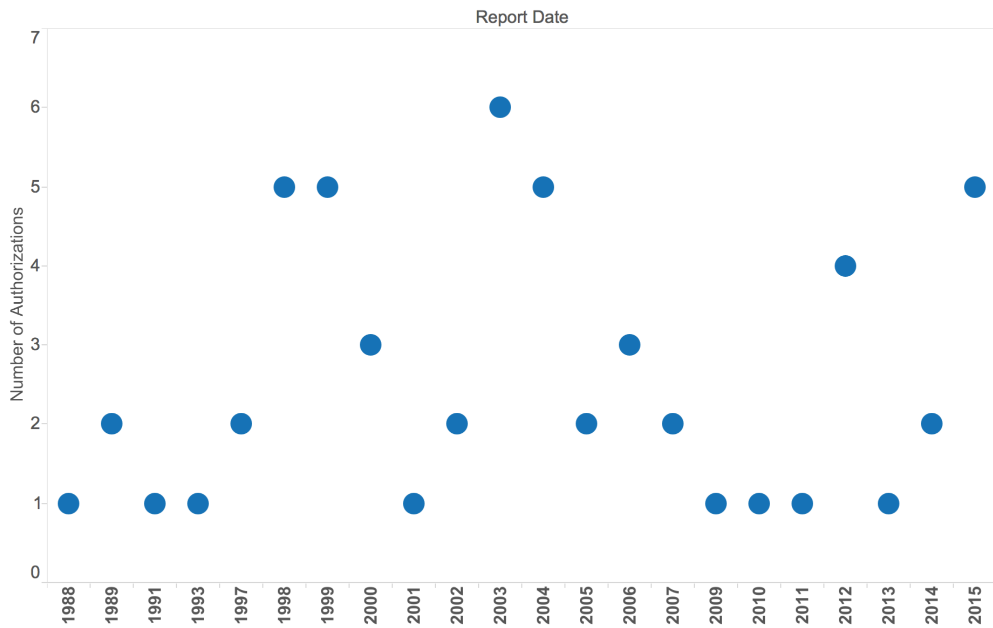


Figure 7. Number of Authorizations for Technical Information Sharing as a Function of Time.

Authorizations as pertaining to technical information sharing remain steady with minor fluctuations. It is unreliable to evaluate trends when there is no statistical analysis; one was not performed for this work and as such, trends discussed above should be considered with reservations.

No significant trends for non-specifically authorized destinations as pertaining to Technical Information Sharing were identified, while China is the dominant country in Engineering Services and Engineering Support. Armenia has a significant number of authorization requests in Engineering Services and India has the second most requests in Engineering Support. The reason for these spikes depends on what is going on in the country at the time of request and thus cannot be fully reconciled.

Records pertaining to sensitive nuclear technologies within the obtained Part 810 General Authorization Data are displayed in Table 3.

Country of Recipient	Type of Transfer	10Cfr810.2 Enrichment	Reprocessing
Canada	Personnel	●	
China	Technical Information Sharing	●	
France	Engineering Services	●	
	Technical Information Sharing	●	
Germany	Technical Information Sharing	●	
Japan	Technical Parts Exchange		●
South Africa	Technical Information Sharing	●	
Taiwan	Technical Information Sharing	●	
United States	Engineering Services		●
	Personnel	●	
	Technical Information Sharing	●	

Table 3. Specific Transfers as Pertaining to Enrichment and Reprocessing to Various Countries.

Data in Table 3 indicates enrichment is the dominant sensitive nuclear technology transfer. It should be noted that all countries receiving these sensitive nuclear technology transfers are considered generally authorized destinations, except for China; the specific transfer was not disclosed in the report. It is recommended that non-specifically authorized destinations should undergo the specific authorization process for sensitive nuclear technology transfers, as countries that are not considered specifically authorized destinations could potentially have inimical interests towards the United States.

To restate, Question 3 asked the following: Are there any patterns in technology being sent to a country e.g. same or different technologies and technology changes over time? China and Argentina were two countries used to explore if there are any patterns in technology being sent overtime. Records were isolated from the Part 810 Excel Spreadsheet and imported into Tableau.

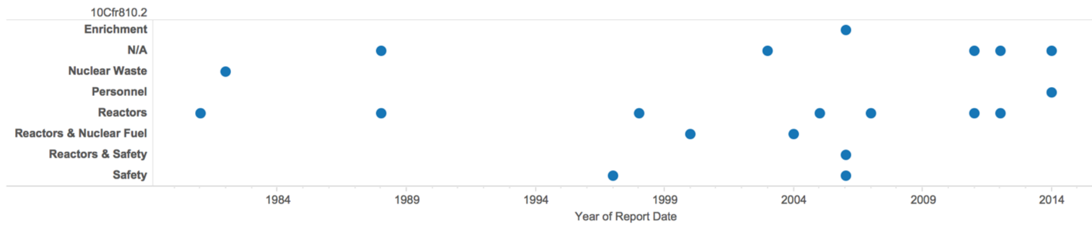


Table 4. Tableau Pictorial Representing the Specific Transfer, as Given by 10CFR810.2 and Report Date for China.

Technology transfers to China, shown in Table 4, follow a pattern for the start-up to decommissioning of a standard nuclear power plant. No specific plant was chosen for reference. The first transfer was a construction & startup audit comparison represented by the Reactors tag under 10 CFR 810.2. Subsequent authorization requests build on the first transfer explained above. For example, transfers related to nuclear waste and reactor safety followed the audit request. Enrichment requests were the last to be observed in 2006. In August 2006, technical information was transferred regarding enrichment; a Chinese delegation visited a U.S. gaseous diffusion plant. This does not mean, however, that any specifications of the gaseous diffusion technology were disclosed. The Plant produced low-enriched uranium, originally as feedstock for military reactors and weapons refining and later for nuclear power fuel; no highly-enriched uranium was produced at this site. The plant has since ceased operations.

	Report Date						
10Cfr810.2	1995	2010	2011	2012	2013	2014	2015
Nuclear Fuel							●
Personnel						●	
Reactors	●		●	●	●	●	●
Reactors & Nuclear Fuel		●					

Table 5. Tableau Pictorial Representing the Specific Transfer, as Given by 10CFR810.2 and Report Date for Argentina.

Argentinian technology transfers revolve around reactor parts and fuel. There are no alarming changes over time, rather there are no transfers related to sensitive nuclear technologies. There are no authorization records prior to 1995, because during the 1960s to the early 1990s Argentina pursued an ambitious program of nuclear energy and technological development, which included construction of an unsafeguarded uranium enrichment facility[28].

Question 4 asks, are there any patterns exhibited by companies in the United States? Does one country always go to the same one or two U.S. companies or do certain companies only transfer to certain countries or regions? Companies with greater than fifteen general authorization transfers were imported into Tableau for more accurate visual analysis, wherein recipient company recurrence was counted. It is important to note that for sensitivity purposes, companies are denoted as U.S. Company 1, U.S. Company 2, etc., while Foreign companies are denoted as Foreign Company 1, Foreign Company 2, etc. There are instances where information is transferred within house to other country subsidiaries, such as U.S. Company 1 (Figure 8), U.S. Company 2 (Figure 9), and U.S. Company 3 (Figure 10).

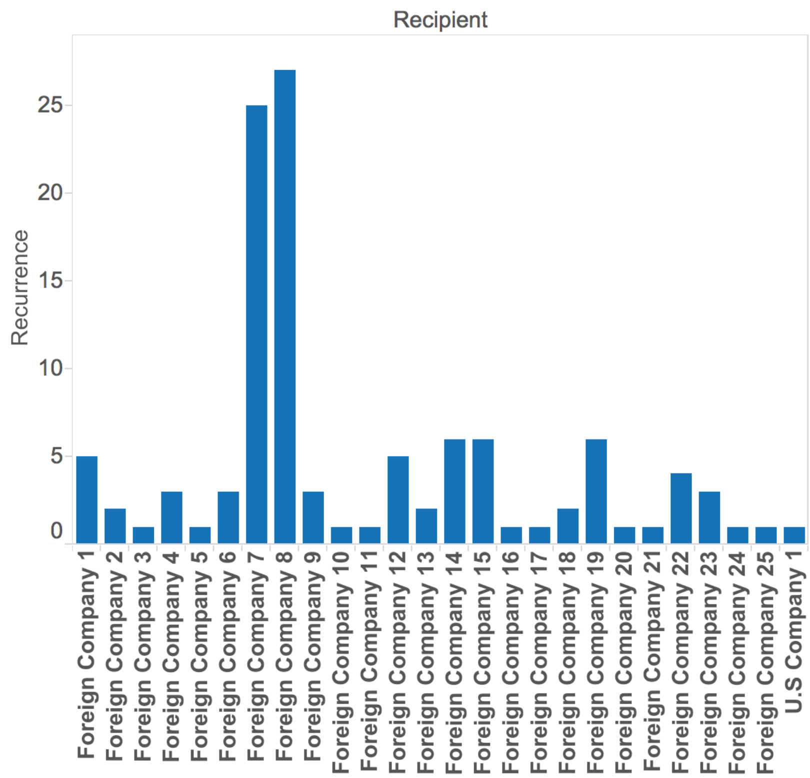


Figure 8. U.S. Company 1 Transfer Recurrence to Various U.S. and Foreign Companies.
Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

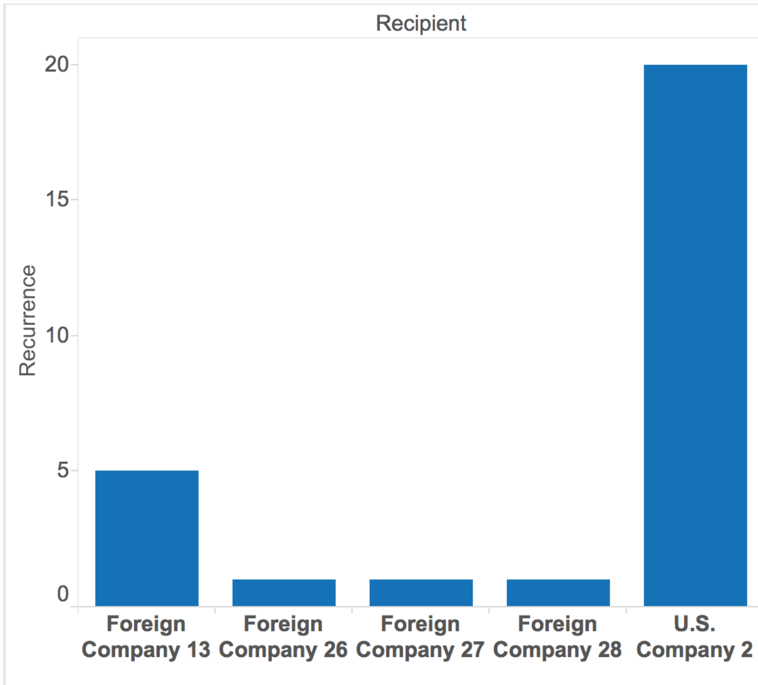


Figure 9. U.S. Company 2 Transfer Recurrence to Various U.S. and Foreign Companies. Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

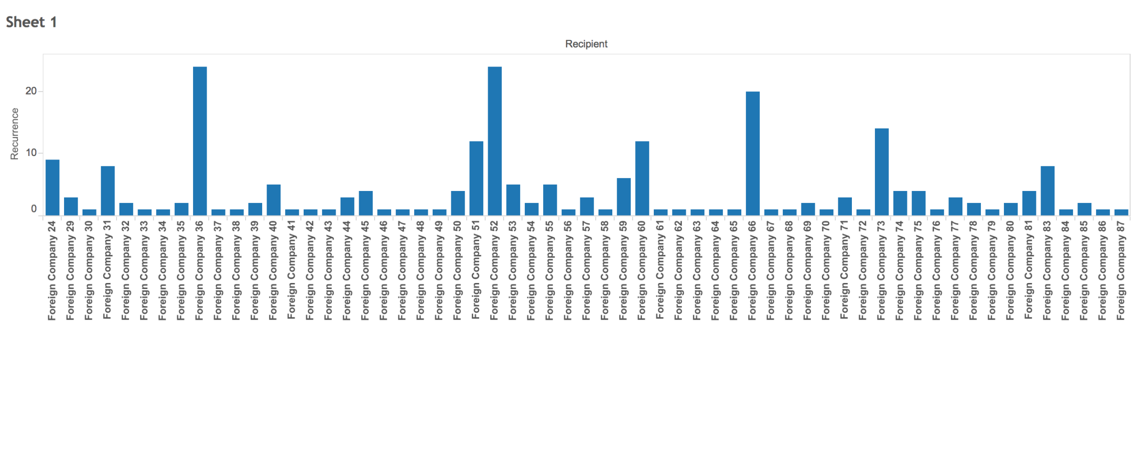


Figure 10. U.S. Company 3 Transfer Recurrence to Various Foreign Companies. Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

Few companies were identified as consistently going to one or two of the same foreign entities. For example, U.S. Company 4 was found to only provide engineering services and supplies to a South Korean company, with twenty-two general authorization requests. Given company isolation, specific transfers were explored to ensure non-inimical intentions and are given in Table 6 below.

	Report Date
Specific Transfer	2015
ALARA Review	●
APR1400 Design Certification	●
Consulting Support Services	●
KJRR Shielding Design & RMS Setpoint Analysis	●
Korean Domestic Nuclear Power Plants Design Evaluation	●
Preparation of NRC RAI Responses	●
Radiation Protection Design	●
Radiation Protection Design Review	●

Table 6. Specific Transfer and Transfer Date to a South Korean Company from U.S. Company 4.

The above table shows no alarming transfers; there is no sensitive nuclear technology, nor dual use items. All requests were made in 2015 and included non-inimical transfers, such as radiation protection designs, reviews, and consulting services.

As shown in Figure 11, U.S. Company 5 has significant transfers to a Canadian company. Company recurrence is counted over all years contained within the Part 810 General Authorization Data (1981-2015). Given Canada is a generally authorized destination and both are power generating stations it is expected that technical information be shared.

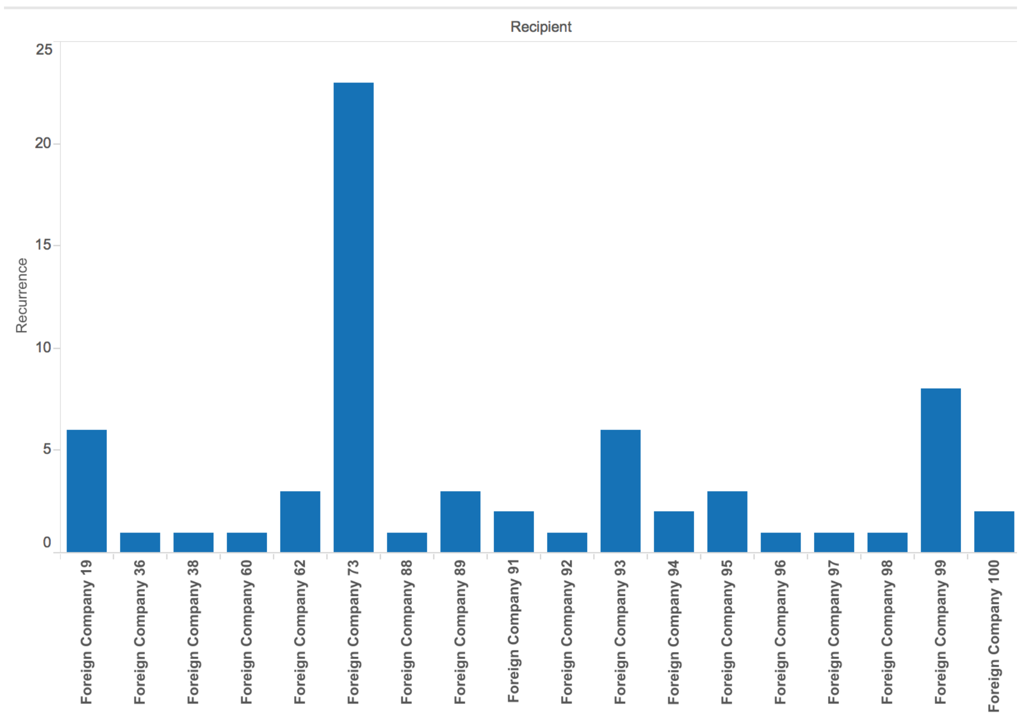


Figure 11. U.S. Company 5 Transfer Recurrence to Various Foreign Companies. Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

U.S. Company 6 and a Swiss company have been identified as having a significant amount of technology transfers, as shown in Figure 12. Company recurrence is once again counted over all years contained within the Part 810 General Authorization Data (1981-2015). It can’t be said with confidence why there are increased transfers between these two companies, as political relationships and turmoil are not known, nor considered as a variable.

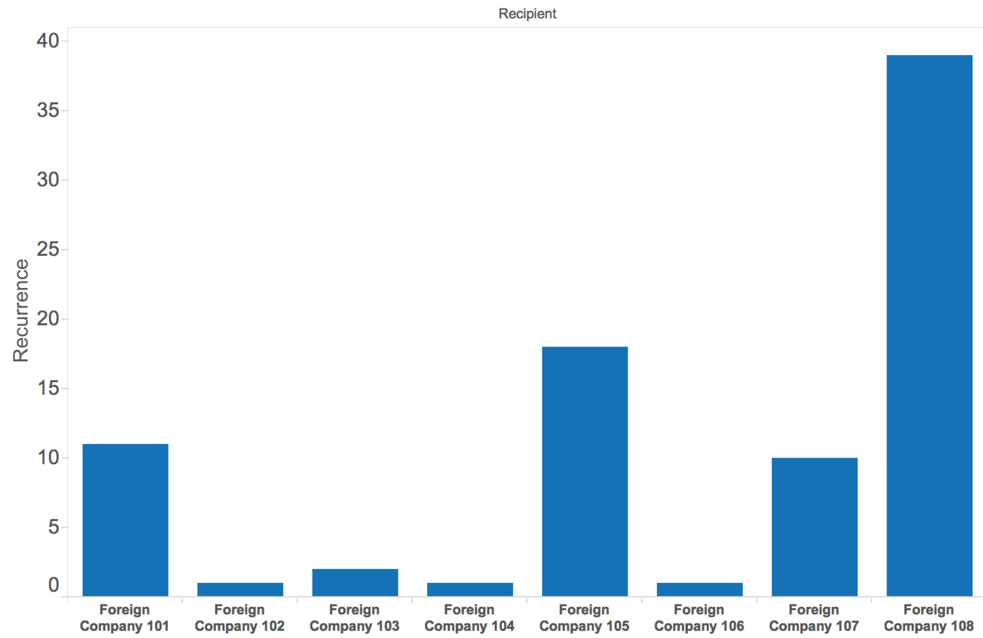


Figure 12. U.S. Company 6 Transfer Recurrence to Various Foreign Companies. Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

This partnership was further explored by identifying specific transfers.

	Type of Transfer			
Specific Transfer	**Engineering Services**	**Engineering Support**	**Personnel**	**Technical Information Sharing**
N/A	●			●
Nuclear Regulatory Safety		●		
Personnel Training			●	
Probabilistic Safety Assessment	●			
Severe Accident Studies, Risk Analysis, & General Safety Analysis		●		
Training			●	

Table 7**.** Specific Transfers Between U.S. Company 6 and a Swiss Company.

All transfers shown pertain to safety and training. Additionally, Switzerland is a generally authorized country, thus the association is favorable.

U.S. Company 7 favors a U.S. power company, as given in Figure 13. The difference between the minimum and maximum recurrence points is only eight. Company 7 only provided a U.S. power company with engineering services for their reactor design.

Figure 13. U.S. Company 7 Transfer Recurrence to Various U.S. and Foreign Companies. Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

U.S. Company 8 provided significant services to a U.S. company as illustrated in Figure 14. Engineering services were carried out within the United States. No technical information was shared.

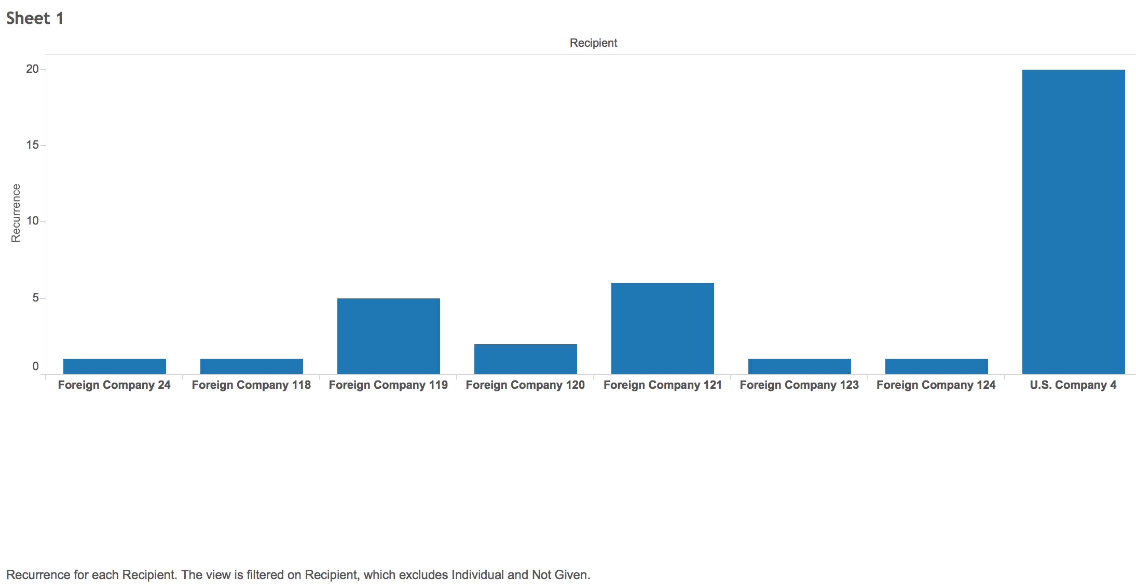


Figure 14. U.S. Company 8 Transfer Recurrence to Various U.S. and Foreign Companies. Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

U.S. Company 9, based out of Oregon, has transferred a significant amount of information to a Swedish company. U.S. Company 9's general authorizations are shown in Figure 15. As in many other cases Sweden is on the general authorization list. Moreover, material sent was either hafnium rods or zircaloy-4 channel strip, essential reactor materials.

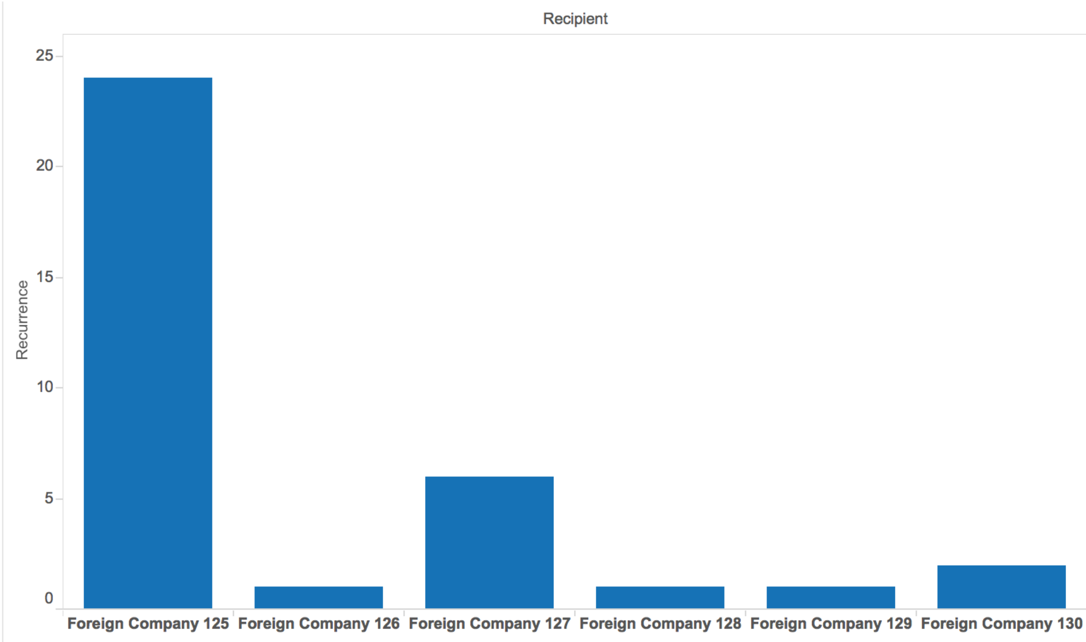


Figure 15. U.S. Company 9 Transfer Recurrence to Various Foreign Companies.
Note: Companies are denoted as Company 1, 2, etc. for sensitivity purposes.

No one country always goes to the same U.S. company; there is a significantly broad spectrum. Certain U.S. companies, on the contrary, do as exhibited in Figures 11-15. Table 8 summarizes part of Question 4, which asks: Do certain companies only transfer to certain countries or regions? Of utmost importance, as stated previously in this section, all recipient countries given are generally authorized destinations, including transfers within the United States.

U.S. Company	Recipient Country
U.S. Company 4	Korea
U.S. Company 5	Canada
U.S. Company 6	Switzerland
U.S. Company 7	United States
U.S. Company 8	United States
U.S. Company 9	Sweden

Table 8. U.S. Companies that Favor One Recipient Country.

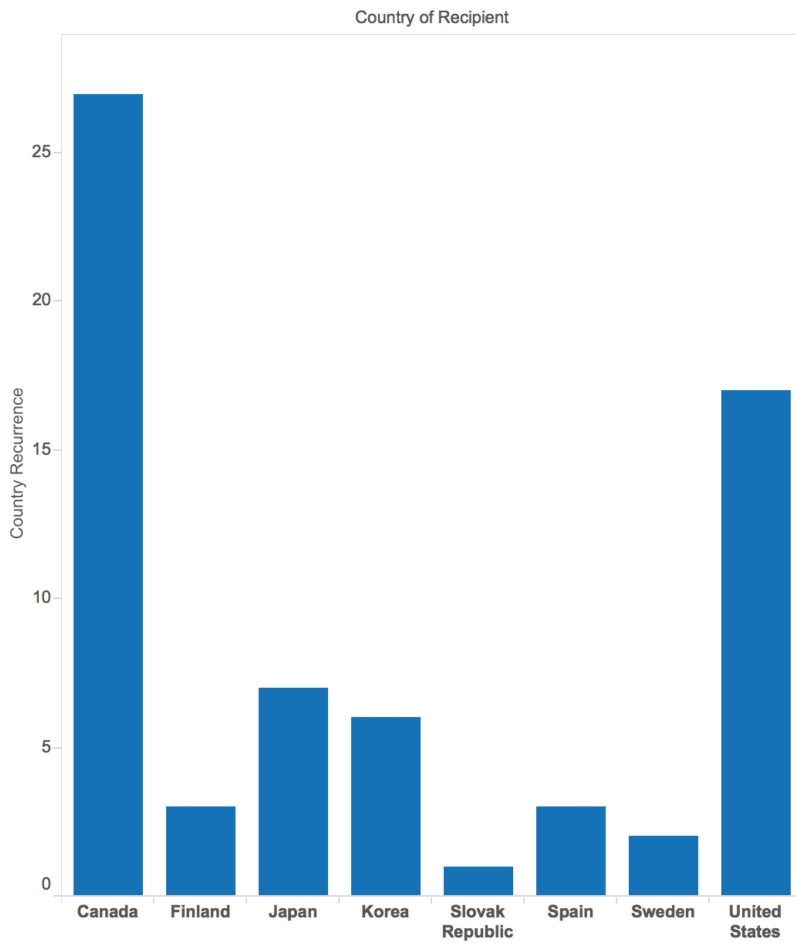


Figure 16. U.S. Company 5 Transfer Recurrence to Specific Foreign Countries.

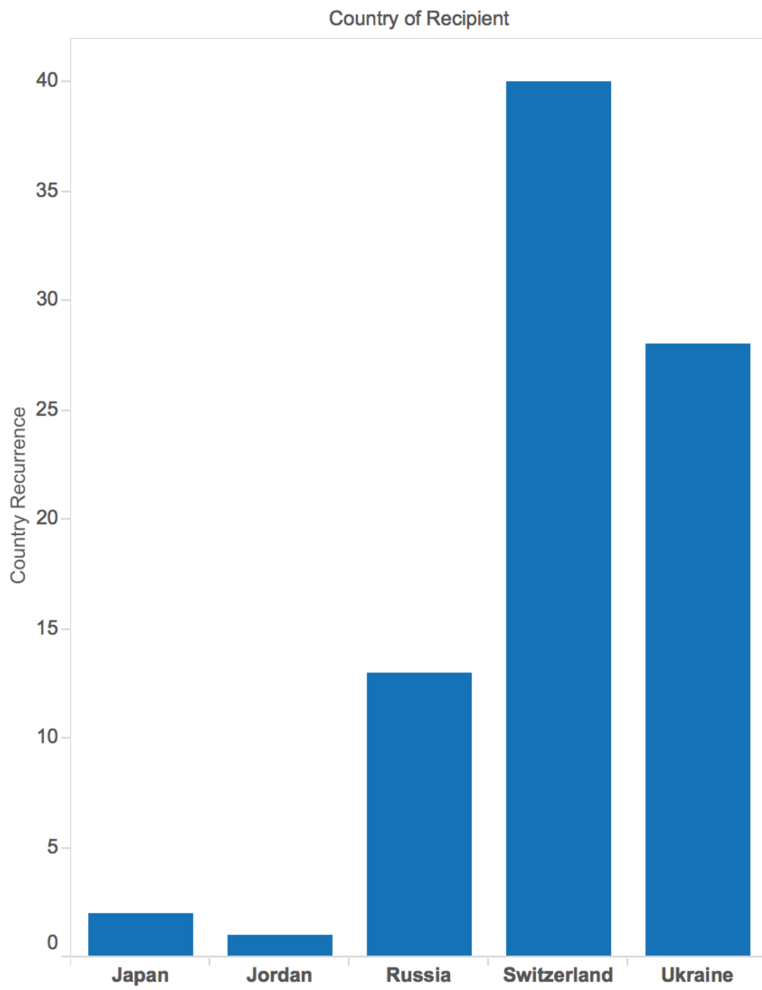


Figure 17. U.S. Company 6 Transfer Recurrence to Specific Foreign Countries.

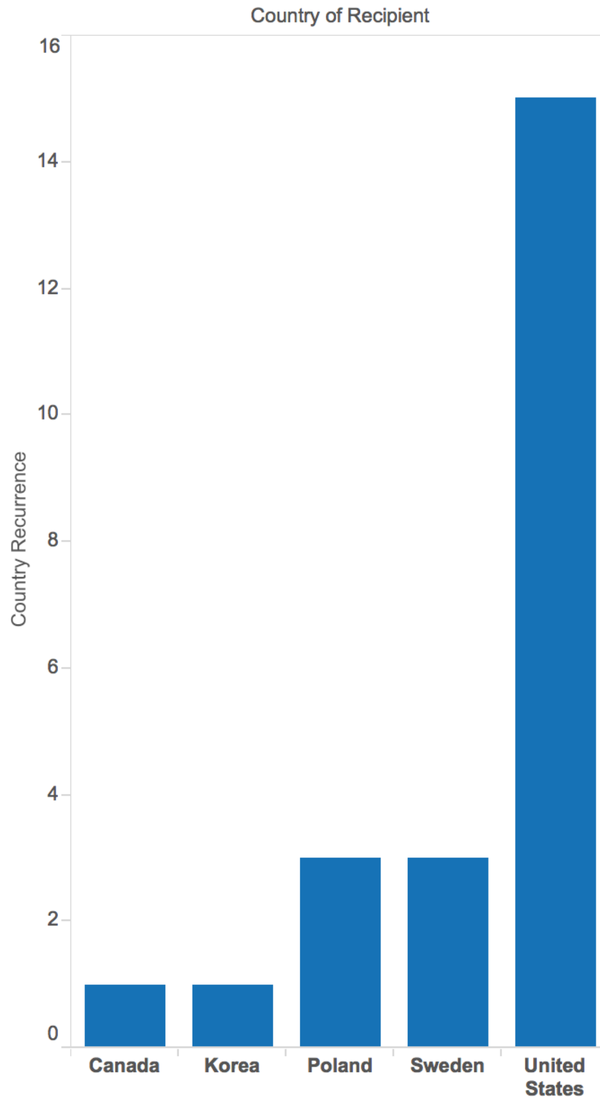


Figure 18. U.S. Company 7 Transfer Recurrence to Specific Foreign Countries.

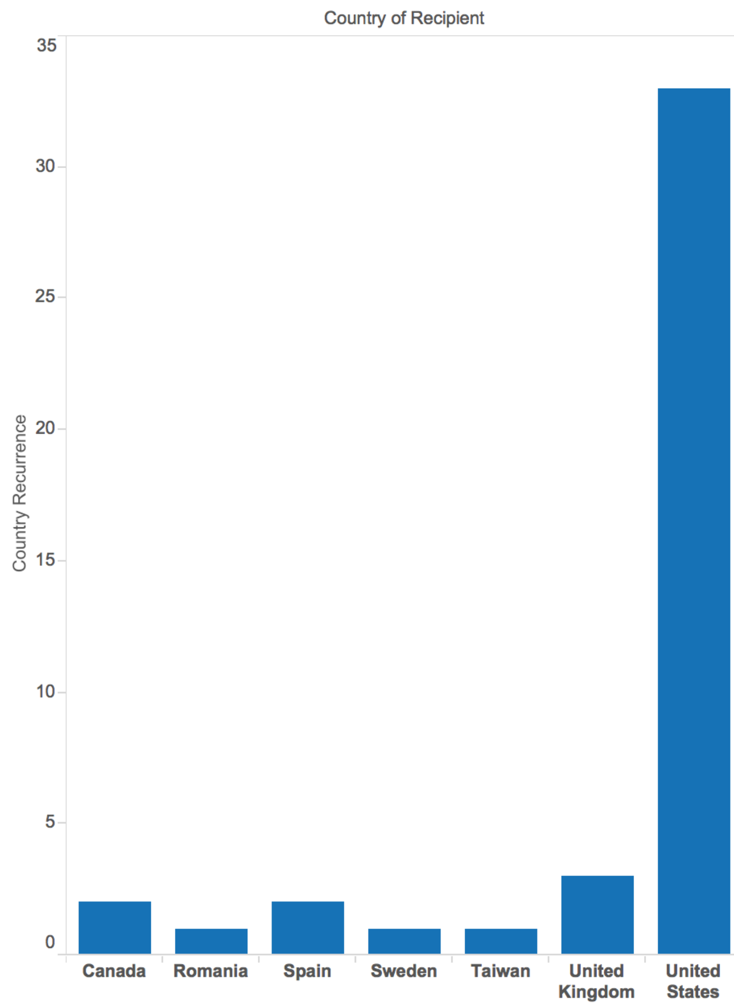


Figure 19. U.S. Company 8 Transfer Recurrence to Specific Foreign Countries.

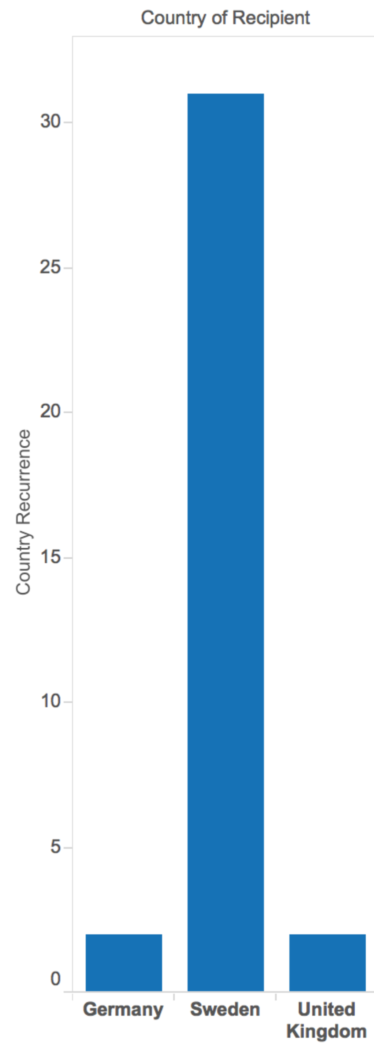


Figure 20. U.S. Company 9 Transfer Recurrence to Specific Foreign Countries.

To reiterate, Figures 16-20 indicate that certain U.S. companies do favor specific countries. These countries, however, are Specifically Authorized Destinations and thus are favorable.

IV. C Prediction: Number of Authorizations to be Received Per Year

Prediction was carried out using Tableau's software package entitled Forecast. Attempts were made to predict how many authorization requests would be received for a given year. Results obtained using this program were also supplemented by simple Excel modeling. Few cases were successfully run, as Tableau's Forecast option is only successful with minimal gaps. The Part 810 General Authorization Data contained gaps or minimal records i.e. a record count below 60 for the following years: 1983-1986, 1989-1994, 1997, 2006, and 2008. Although excel modeling proved lucrative, it does not carry any argumentative weight, as variables, such as market conditions were not considered. A country's specific needs were not known. The excel modeling is only given as an extremely vague estimate.

Data obtained for Slovenia was used as a test case. It is important to note that in Figure 21, the shaded region and line contained therein are the forecast or prediction for years to come. Figure 21 was contrived using Tableau. The spike in 2015 was a result of increased funding, as categorized under Engineering Support.

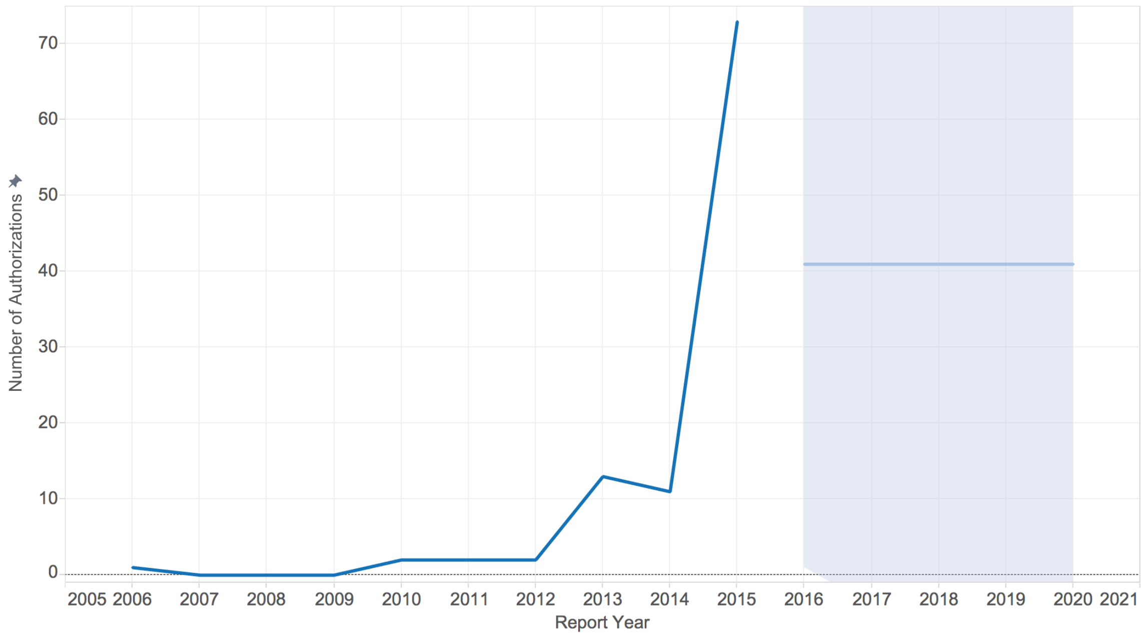


Figure 21. Slovenia Forecast as Given by Tableau.

As shown, the program predicts that for coming years the likely number of reports to be received is approximately 41, but it can range between 0 and 75; an extremely large spread. Data was subsequently plotted in excel using a simple linear regression and exponential fit. It is important to note that rather than using the Report Years themselves, a number equivalent was assigned to each. For example, year 2006 corresponds to 1, while 2007 corresponds to 2.

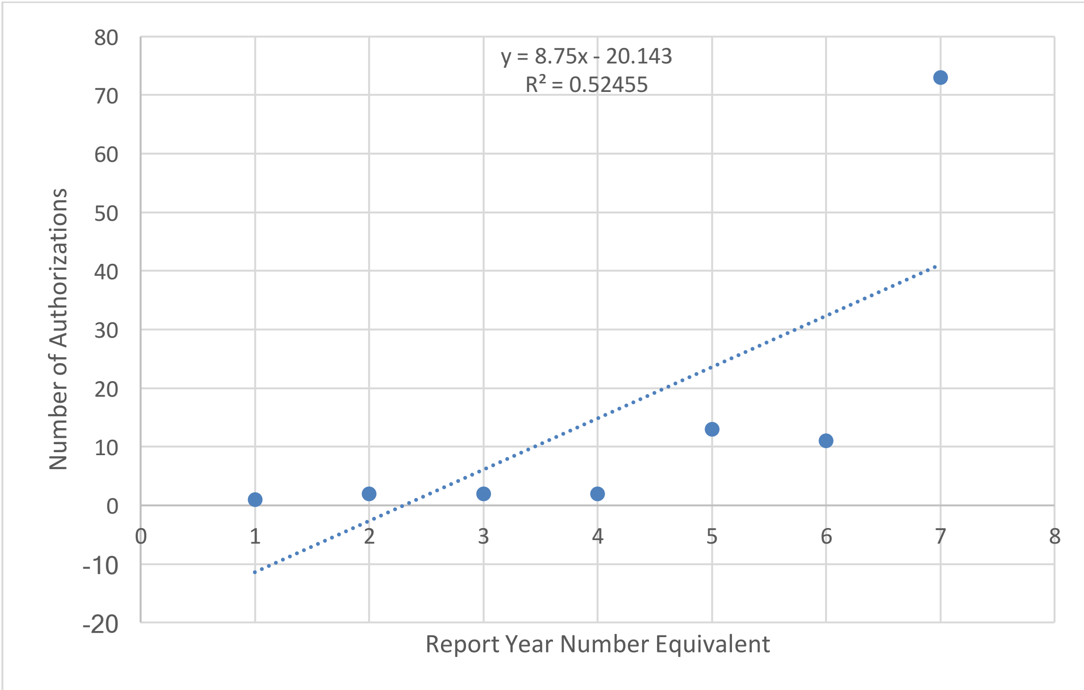


Figure 22. Linear Fit of the Number of Received Authorizations as a Function of Report Year for Slovenia.

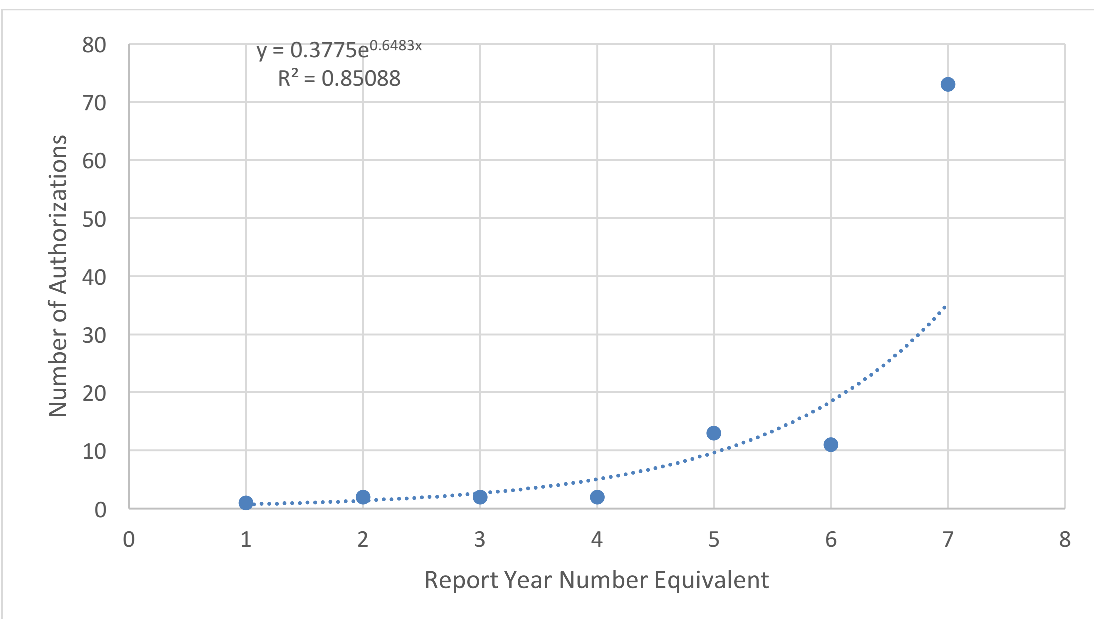


Figure 23. Exponential Fit of the Number of Received Authorizations as a Function of Report Year for Slovenia.

The R^2 value is known as the coefficient of determination. This value is interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variable[29]. The value of the coefficient of determination as 1 indicates the dependent variable can be predicted without error from the independent variable[29]. The R^2 value for the linear and exponential fits are 0.53 and 0.85 respectively.

Equations given in Figures 22 and 23 were used to predict the number of authorizations to be received in 2019, to compare against Tableau's forecast. The number of authorizations, as determined by the linear fit are 50, while the exponential fit yields 66 authorizations. Output of the linear fit falls within Tableau's forecast. Percent error between the linear fit result and Tableau's predicted 41 authorizations was determined by Equation 1 and is found to be 22 %, while error in the exponential prediction is 66 %.

$$\%\ \text{error} = \left|\frac{\text{experimental} - \text{theoretical}}{\text{theoretical}}\right| \qquad \text{(Equation 1)}$$ [30]

Analysis was carried out for Argentina, however, given excel modeling uncertainty, it is not included.

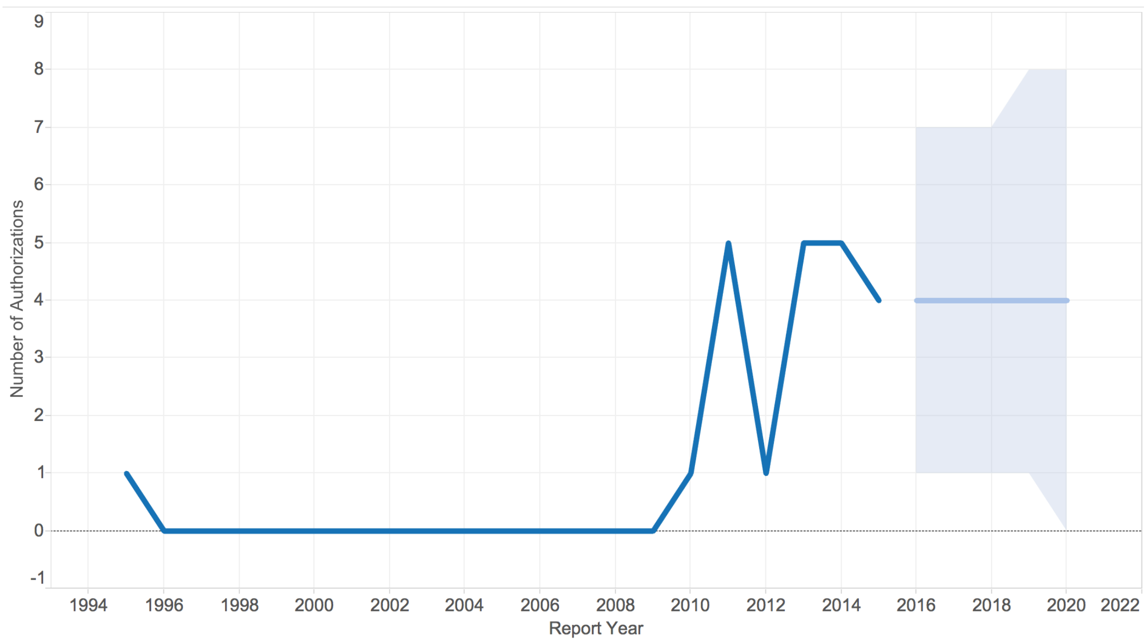


Figure 24. Argentinian Forecast as given by Tableau.

Tableau predicts the number of authorizations to be received ranges between 1 and 7 from 2016 to 2018, while the predicted records for 2019 and 2020 are between 0 and 8. Contained within the range, is the horizontal line, representing the specific number of predicted authorizations to be received (4 authorizations).

Tableau's forecast can be used as an estimate to predict when a general authorization is to be received given a year. Excel modeling, on the contrary, should be used sparingly, as unknown variables, such as country needs, are not taken into consideration.

IV.D News Sources

Maltego Classic Edition (CE) was the primary tool used for predicting when a General Authorization is to be expected. The program was set up such that it will search for a specific country and tag as given in Excel Spreadsheet Categorization. Countries to be explored include those destined as generally authorized and participants in The Nuclear Suppliers Group (NSG), as given in Appendix I. U.S. Companies identified in Answered Questions to be Presented to DOE were also examined. An example phrase search would be “Argentina” “Uranium Conversion”, while U.S. companies were searched for on an individual basis. It is important to note that all keywords used for tagging cannot be searched for at the same time, as then each phrase will need to be found in the same source.

Upon entering the phrase of choice, Maltego CE will generate a graph showing link analysis between all phrases, in addition to a report that will give URL links, as well as snippets from the source of interest. It is up to the discretion of the user to determine which sources are viable and useful. Example nodular graphs for Slovenia and Australia are given in Figures 25 and 26, while report fragments are shown in Figures 27 and 28, as reports can span 60+ pages. It is important to note that the code searches all sources; there is no reason Wikipedia sources are only given.

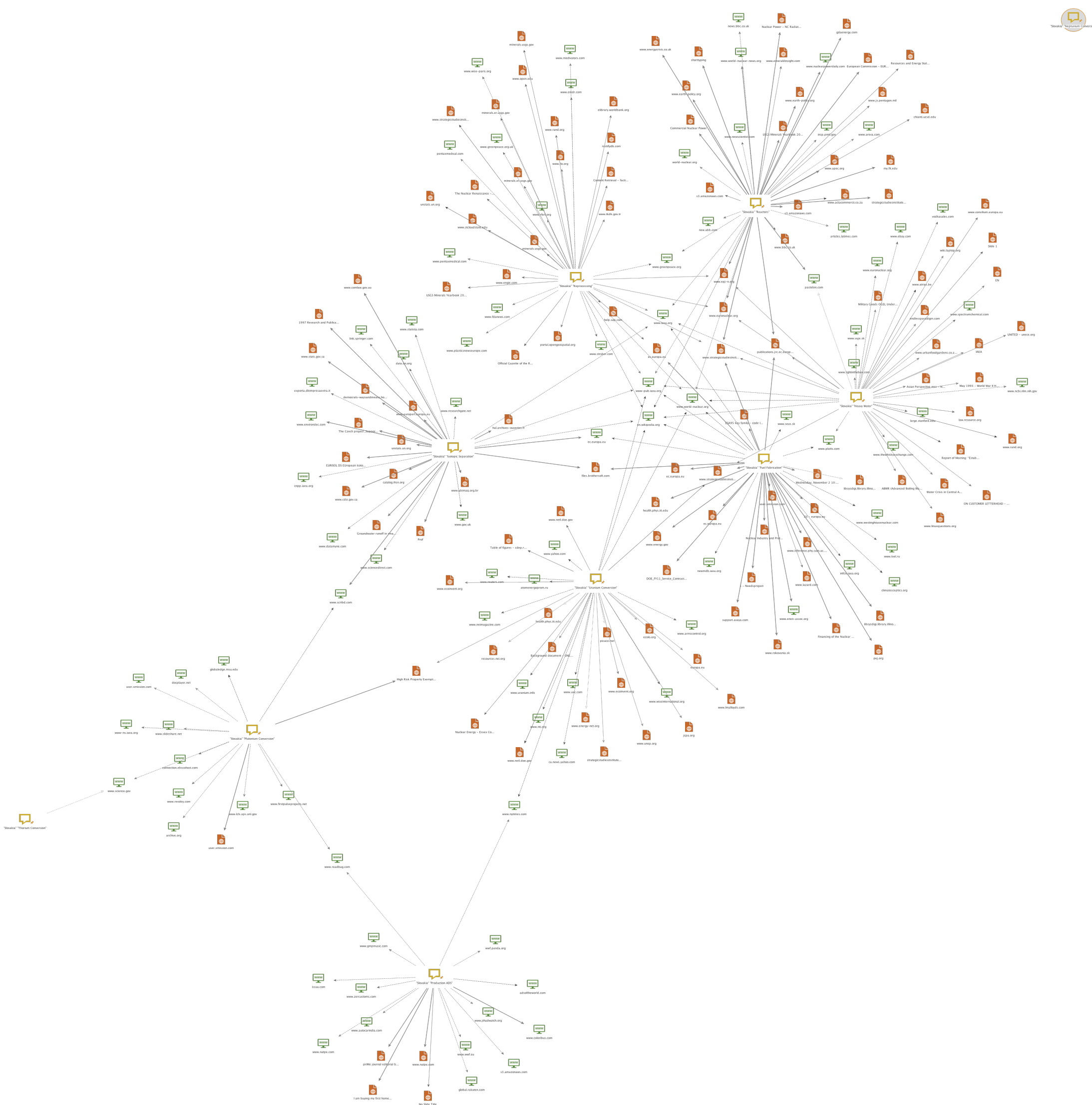


Figure 25. Example Nodular Graph Generated using Maltego CE for Slovenia.

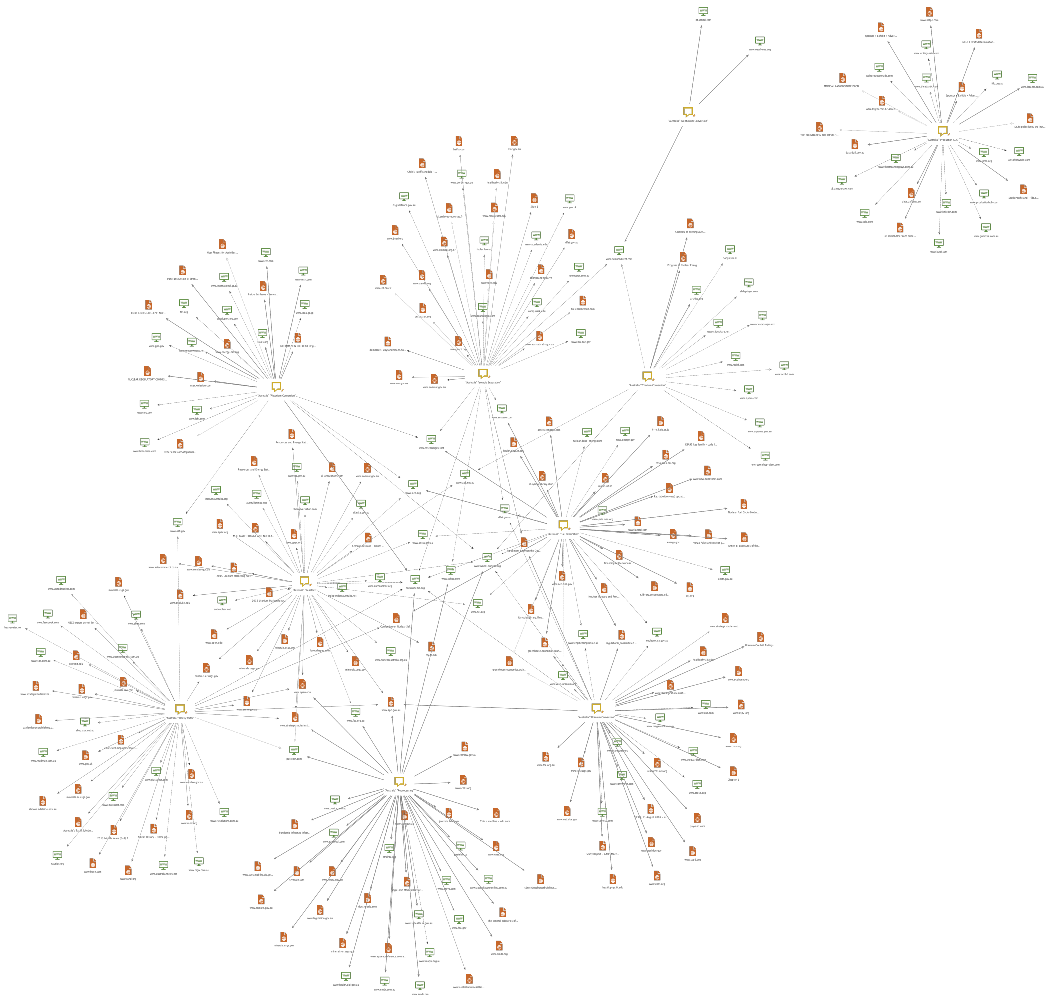


Figure 26. Example Nodular Graph Generated using Maltego CE for Australia.

Snippet(s):

Question: ["Slovenia" "Fuel Fabrication"] using bing

Nuclear program of Iran - Wikipedia

[https://en.wikipedia.org/wiki/Nuclear_program_of_Iran]

2 pellet fuel fabrication and a chemical department whose goal was the conversion of U 3 O 8 to nuclear grade UO 2.

Question: ["Slovenia" "Fuel Fabrication"] using bing

List of companies in the nuclear sector - Wikipedia, the ...

[https://en.wikipedia.org/wiki/List_of_companies_in_the_nuclear_sector]

List of companies in the nuclear sector This article has multiple issues. ... Slovenia: Electricity generation: GEN Energija, a state-owned company, ...

Snippet(s):

Question: ["Slovenia" "Isotopic Separation"] using bing

Red mercury - Wikipedia, the free encyclopedia

[https://en.wikipedia.org/wiki/Red_Mercury]

Red mercury is a hoax substance of uncertain composition purportedly used in the creation of nuclear bombs, as well as a variety of unrelated weapons systems.

Question: ["Slovenia" "Isotopic Separation"] using bing

List of nuclear reactors - Wikipedia

[https://en.wikipedia.org/wiki/List_of_nuclear_reactors]

National Institute for Research and Isotopic Separation, Govora, 170 km west of Bucharest - no research reactors, but instead devoted to heavy water production;

Snippet(s):

Question: ["Slovenia" "Uranium Conversion"] using bing

Nuclear power in Australia - Wikipedia, the free encyclopedia

[https://en.wikipedia.org/wiki/Nuclear_power_in_Australia]

Nuclear power plant concepts and proposals 1952 Upper Spencer Gulf, South Australia. In 1952, South Australian Premier Thomas Playford expressed with confidence that ...

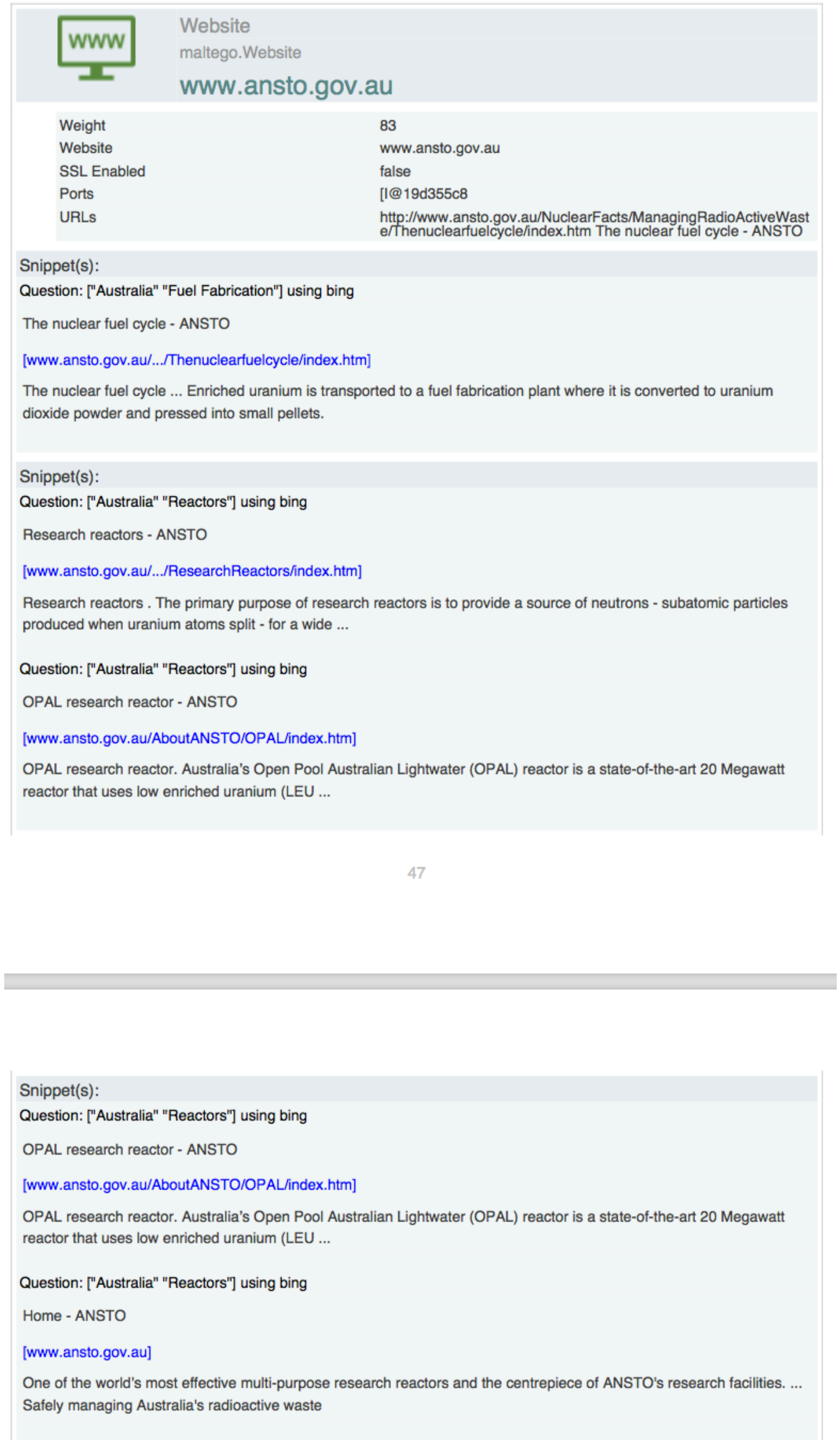
Question: ["Slovenia" "Uranium Conversion"] using bing

List of companies in the nuclear sector - Wikipedia, the ...

[https://en.wikipedia.org/wiki/List_of_companies_in_the_nuclear_sector]

List of companies in the nuclear sector This article has multiple issues. Please help improve it or ... Slovenia: Electricity generation: GEN Energija, ...

Figure 27. Example Fragment from the Slovenia Report.


Website
maltego.Website
www.ansto.gov.au

Weight	83
Website	www.ansto.gov.au
SSL Enabled	false
Ports	[I@19d355c8
URLs	http://www.ansto.gov.au/NuclearFacts/ManagingRadioActiveWaste/Thenuclearfuelcycle/index.htm The nuclear fuel cycle - ANSTO

Snippet(s):

Question: ["Australia" "Fuel Fabrication"] using bing

The nuclear fuel cycle - ANSTO

[www.ansto.gov.au/.../Thenuclearfuelcycle/index.htm]

The nuclear fuel cycle ... Enriched uranium is transported to a fuel fabrication plant where it is converted to uranium dioxide powder and pressed into small pellets.

Snippet(s):

Question: ["Australia" "Reactors"] using bing

Research reactors - ANSTO

[www.ansto.gov.au/.../ResearchReactors/index.htm]

Research reactors . The primary purpose of research reactors is to provide a source of neutrons - subatomic particles produced when uranium atoms split - for a wide ...

Question: ["Australia" "Reactors"] using bing

OPAL research reactor - ANSTO

[www.ansto.gov.au/AboutANSTO/OPAL/index.htm]

OPAL research reactor. Australia's Open Pool Australian Lightwater (OPAL) reactor is a state-of-the-art 20 Megawatt reactor that uses low enriched uranium (LEU ...

47

Snippet(s):

Question: ["Australia" "Reactors"] using bing

OPAL research reactor - ANSTO

[www.ansto.gov.au/AboutANSTO/OPAL/index.htm]

OPAL research reactor. Australia's Open Pool Australian Lightwater (OPAL) reactor is a state-of-the-art 20 Megawatt reactor that uses low enriched uranium (LEU ...

Question: ["Australia" "Reactors"] using bing

Home - ANSTO

[www.ansto.gov.au]

One of the world's most effective multi-purpose research reactors and the centrepiece of ANSTO's research facilities. ... Safely managing Australia's radioactive waste

Figure 28. Example Fragment from the Australia Report.

IV.E Data Mining and Patterns

The Apriori Algorithm was used for data mining purposes in addition to identifying patterns that in turn resulted in association mining rules. Sample output is given in Figure 29 and explained upon below.

```
=== Run information ===

Scheme:     weka.associations.Apriori -N 10 -T 0 -C 0.4 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:   Argentina-weka.filters.unsupervised.attribute.Remove-R1
Instances:  22
Attributes: 5
            Report Year
            Type of Transfer
            10CFR810.2
            Company
            Recipient
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.25 (5 instances)
Minimum metric <confidence>: 0.4
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 9

Size of set of large itemsets L(3): 1

Best rules found:

 1. Type of Transfer=Technical Information Sharing 7 ==> Recipient=[redacted] 7    <conf:(1)>
lift:(1.57) lev:(0.12) [2] conv:(2.55)
 2. Type of Transfer=Engineering Services 6 ==> 10CFR810.2=Reactors 6    <conf:(1)> lift:(1.38)
lev:(0.07) [1] conv:(1.64)
 3. Report Year=MMXIII 5 ==> Recipient=[redacted] 5    <conf:(1)> lift:(1.57) lev:(0.08) [1]
conv:(1.82)
 4. Company =[redacted]
5 ==> Recipient=[redacted] 5    <conf:(1)> lift:(1.57) lev:(0.08) [1] conv:(1.82)
 5. Type of Transfer=Engineering Support Recipient=[redacted] 5 ==> 10CFR810.2=Reactors 5
<conf:(1)> lift:(1.38) lev:(0.06) [1] conv:(1.36)
 6. Type of Transfer=Engineering Support 10CFR810.2=Reactors 6 ==> Recipient=[redacted] 5
<conf:(0.83)> lift:(1.31) lev:(0.05) [1] conv:(1.09)

 7. Recipient=[redacted] 14 ==> 10CFR810.2=Reactors 11    <conf:(0.79)> lift:(1.08) lev:(0.04) [0]
conv:(0.95)
 8. Company =[redacted] 8 ==> 10CFR810.2=Reactors 6    <conf:(0.75)> lift:(1.03) lev:(0.01)
[0] conv:(0.73)
 9. Company =[redacted] 8 ==> Recipient=[redacted] 6    <conf:(0.75)> lift:(1.18) lev:(0.04) [0]
conv:(0.97)
10. 10CFR810.2=Reactors 16 ==> Recipient=[redacted] 11    <conf:(0.69)> lift:(1.08) lev:(0.04) [0]
conv:(0.97)
```

Figure 29. Sample Apriori Output for Argentina.
Note: The company name is blacked out for sensitivity purposes.

Apriori, as used via Weka, was set up such that it would find association rules for data mining using categories in the constructed Excel Spreadsheet. The chosen variables were Report Year, Type of Transfer, 10 CFR 810.2, Company, and Recipient. It is important to note that these variables are not numeric, but rather categorical. For this reason, the Report Year is presented in roman numerals, rather than given as a discrete number.

The run information, as given above, will present the following:

- Scheme for Learning Association used: Apriori
- Relation Name: Argentina
- Number of Instances in the Relation: 22
- Number of Attributes in the Relation: 5

The program initiates by generating the sets of large itemsets found for each support size that is considered. For example, in the case above, 10 itemsets were found to have the required minimum support, as given by Size of set of large itemsets L(1):10. Apriori, by default, will produce 10 rules. It will begin with a minimum support of 100 % of the data items and subsequently decrease in 5% increments until there are a minimum of ten rules with required confidence. Minimum confidence is 0.4 or 40 %; minimum support decreased to 0.25 or 25 % before the required number of rules were generated. Rule generation required 15 cycles. Best rules found are given at the end of the run information. The number that proceeds ==> is indicative of the number of items covered by the premise. The numeric following, gives the number of those items for which the rule's consequent holds[31]. For example, take the first rule given in Figure 29; If the Type of Transfer is technical information sharing, it can be said with 100 % confidence that the likely recipient will be a specific Argentinian company. Rules such as these, can be applied to how data is mined, as well as predicting what could be seen in the future. Apriori was applied to all countries in the NSG.

CHAPTER V

CONCLUSIONS

Data objectives were carried out and completed in conjunction with ANL. The constructed Excel spreadsheet enables ad hoc searching, while the analytic software Tableau was used to answer questions to be presented to the DOE. No significant trends for non-specifically authorized destinations as pertaining to Technical Information Sharing were identified, while China is the dominant country in Engineering Services and Engineering Support. Armenia has a significant number of authorization requests in Engineering Services and India has the second most requests in Engineering Support. The reason for these spikes depends on what is going on in the country at the time of request and thus cannot be fully reconciled. Additionally, the data package can be used to determine what type of reporting is necessary and customize accordingly. For example, as previously stated, non-specifically authorized countries that exceed a certain threshold of general authorizations should be supervised to ensure that these countries of interest are not using multiple general authorizations to circumvent specific authorization. The Apriori algorithm was implemented to recognize patterns and establish association rules, such that patterns and ad hoc searching can be modified accordingly. Apriori was run for all countries in the NSG. Apriori could also be used for prediction. For example, if given the Type of Transfer it can be predicted with a certain confidence that one given company is the likely recipient. Tableau's interim program, Forecast, has been used in conjunction with Excel modeling to predict when a General Authorization is to be expected. Tableau's forecast can be used as an estimate to predict when a general authorization is to be received given a year. Excel modeling, on the contrary, should be used sparingly, as unknown variables, such as country needs, are not taken into consideration. Maltego CE was the primary tool used for predicting when a General Authorization is to be expected. The program was set up such that it will search for a specific country and tag as given in Excel Spreadsheet Categorization. Upon entering the phrase of choice, Maltego CE will generate a graph showing link analysis between all phrases, in addition to a report that will give URL links, as well as snippets from the source of interest. It is up to the discretion of the user to determine which sources are useful. The tools created and utilized above have been tested and prove to be a reliable method for monitoring and analyzing the Part 810 General Authorization Data.

REFERENCES

1. "10 CFR Part 810." National Nuclear Security Administration. Web. 14 Sept. 2016, from https://nnsa.energy.gov/aboutus/ourprograms/nonproliferation-0/npac/policy/10cfr810

2. Company files for Part 810 authorization with DOE. Retrieved December 23, 2016, from https://business.usa.gov/export-portal?civilnuclear/eg_main_022102.asp

3. Elmore, R. A., & Charlton, W.S. (2014) Nuclear Nonproliferation Analysis Using Agent based Modeling in an Entropy Empowered Intelligent Agent Bayesian Framework. Retrieved February 16, 2017, from http://informs-sim.org/wsc14papers/includes/files/363.pdf

4. Listner, C., Canty, M.J., Stein, G., Rezniczek, A., & Niemeyer, I. (2014). Quantifying Detection Probabilities for Proliferation Activities in Undeclared Facilities. Retrieved February 17, 2017, from https://www.iaea.org/safeguards/symposium/2014/home/eproceedings/sg2014-papers/000303.pdf

5. Renda, G., Kim, L., Jungwirth, R., Pabian, F., Wolfart, E., & Cojazzi, G.G.M. (Mar 2015). *The Potential of Open Source Information in Supporting Acquisition Pathway Analysis to Design IAEA State Level Approaches* (IAEA-CN--220). International Atomic Energy Agency (IAEA)

6. Zentner, M. D., Coles, G. L., & Talbert, R. J. (2005, September). Nuclear Proliferation Technology Trends Analysis. Retried February 17, 2017, from http://fissilematerials.org/library/zen05.pdf

7. Charlton, W.S. & Ford, D.G, "Assessment Tool for Nuclear Weapon Acquisition Pathways," Annual Meeting of the Institute for Nuclear Materials Management, July 8-12, 2007, Tucson, AZ.

8. Hawila, Mohammad A. and Chirayath, Sunil S. (2016) "Nuclear Security Risk Analysis: An Insider-Outsider Collusion Scenario," *International Journal of Nuclear Security*: Vol. 2: No. 2, Article 7.

9. Reed, B.K. (1994). Models for proliferation interdiction response analysis. (Unpublished master’s thesis). Naval Postgraduate School, September. Retrieved February 17, 2017, from

http://calhoun.nps.edu/bitstream/handle/10945/43015/94Sep_Reed.pdf?sequence=1

10. Tarnopol, R. (2014, April 15). How to OCR Documents for Free in Google Drive. Retrieved January 5, 2016, from http://computers.tutsplus.com/tutorials/how-to-ocr-documents-for-free-in-google-drive--cms-20460

11. Python-based tools for document analysis and OCR. Retrieved January 5, 2016, from https://github.com/tmbdev/ocropy

12. Which OCR Engine is better: Tesseract or OCRopus? (2012, April 6). Retrieved January 4, 2016, from http://stackoverflow.com/questions/10032923/which-ocr-engine-is-better-tesseract-or-ocropus

13. OCR. (2015, March 31). Retrieved January 4, 2016, from https://help.ubuntu.com/community/OCR

14. Thompson, S., Weimar, M., Daly, D., Sanfilippo, A., Holder, L., Whitney, PD., Wood, TW., Brothers, AJ., Cook, D. (2011, September 1). Model and Analytic Processes for Export License Assessments. Retrieved January 4, 2016, from http://www.pnnl.gov/main/publications/external/technical_reports/PNNL-18189.pdf

15. Coles, G., Brothers, A., Olson, J., & Whitney, P. Assessing State Nuclear Weapons Proliferation: Using Bayesian Network Analysis of Social Factors. Pacific Northwest International Conference.

16. Gillen, D. (2014, May 13). Information Analysis Technologies, Techniques and Methods for Safeguards, Nonproliferation and Arms Control Verification Workshop. Lecture presented at Workshop Proceedings in Portland Marriott Downtown Waterfront Hotel, Portland.

17. Collier, N. (2013, August 13). Repast HPC Manual. Retrieved January 4, 2016, from http://repast.sourceforge.net/docs/repast_hpc.pdf

18. Collier, N., & North, M. (2012). Parallel agent-based simulation with Repast for High Performance Computing. Simulation, 89(10), 1215-1235. doi:10.1177/0037549712462620

19. Tableau Software. Retrieved September 4, 2016, from https://en.wikipedia.org/wiki/Tableau_Software#cite_note-7

20. Gondaliya, T., Joshi, Hi., Joshi Ha. (2015, March). New Big Things in Era of Digital Data: 'Big Data'& Big Data Challenges with its Solution Using Different Tools. Paper presented at 10th International Conference CALIBER 2015: At HP University, Shimla. DOI: 10.13140/RG.2.1.3557.0728

21. Lumify Documentation. Retrieved October 10, 2016, from http://lumify.io/docs/

22. Ljubljana, B. L. Orange Data Mining. Retrieved October 11, 2016, from http://orange.biolab.si/

23. Demsar, J. (2013). Orange: Data Mining Toolbox in Python. Journal of machine learning research, 14, 2349-2353.

24. Maltego Classic. Retrieved October 12, 2016, from https://www.paterva.com/web7/buy/maltego-clients/maltego.php

25. Weka 3: Data Mining Software I Java. Retrieved November 21, 2016, from http://www.cs.waikato.ac.nz/~ml/weka/

26. Steinbach, K. (2004, April 18). Data Mining Association Analysis: Basic Concepts and Algorithms Retrieved December 9, 2016

27. International School of Engineering Follow. (2014, June 19). Apriori Algorithm. Retrieved December 09, 2016, from http://www.slideshare.net/INSOFE/apriori-algorithm-36054672

28. Argentina. Retrieved December 13, 2016, from http://www.nti.org/learn/countries/argentina/

29. Coefficient of Determination. Retrieved December 15, 2016, from http://www.bing.com/cr?IG=D1E6D4764DF4418C82EBB2F0B60A0031&CID=3E9990519B566D72305399BD9A676CFA&rd=1&h=1_T5L6ihPFSo2Gx4tops43oMRtY3LqiNwcvTImNFQ1Y&v=1&r=http%3a%2f%2fstattrek.com%2fstatistics%2fdictionary.aspx%3fdefinition%3dcoefficient_of_determination&p=DevEx,5083.1

30. Percent Error Formula. Retrieved December 15, 2016, from http://astro.physics.uiowa.edu/ITU/glossary/percent-error-formula/

31. Aksenova, S.S. (2004). Machine Learning with WEKA WEKA Explorer Tutorial for WEKA Version 3.4.3. Retrieved January 12, 2017, from http://people.sabanciuniv.edu/berrin/cs512/lectures/WEKA/WEKA%20Explorer%20Tutorial-REFERENCE.pdf

32. "Participants." Nuclear Suppliers Group. Web. 06 Oct. 2016.

APPENDIX I

Argentina	China	Germany	Latvia	Poland	South Korea
Australia	Croatia	Greece	Lithuania	Portugal	Spain
Austria	Cyprus	Hungary	Luxembourg	Romania	Sweden
Belarus	Czech Republic	Iceland	Malta	Russia	Switzerland
Belgium	Denmark	Ireland	Mexico	Serbia	Turkey
Brazil	Estonia	Italy	Netherlands	Slovakia	Ukraine
Bulgaria	Finland	Japan	New Zealand	Slovenia	United Kingdom
Canada	France	Kazakhstan	Norway	South Africa	United States

Table 9. Nuclear Suppliers Group Participants[32]

APPENDIX II

ABM	Agent Based Modeling
ANL	Argonne National Laboratory
CFR	Code of Federal Regulations
DOE	Department of Energy
HPC	High Performance Computing
MPI	Message Passing Interface
NNSA	National Nuclear Security Administration
NSG	Nuclear Suppliers Group
NSSPI	Nuclear Security Science and Policy Institute
OCR	Optical Character Recognition
PNNL	Pacific Northwest National Laboratory
SD	System Dynamics
VizQL	Visual Query Language