PNNL-24183

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Development of a SPARK Training Dataset

Task 2 Deliverable

March 2015

AM Sayre
JR Olson

U.S. DEPARTMENT OF
ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.

(8/2010)

PNNL-24183

Development of a SPARK Training Dataset

Task 2 Deliverable

AM Sayre
JR Olson

March 2015

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Executive Summary

In its first five years, the National Nuclear Security Administration's (NNSA) Next Generation Safeguards Initiative (NGSI) sponsored more than 400 undergraduate, graduate, and post-doctoral students in internships and research positions (Wyse 2012). In the past seven years, the NGSI program has, and continues to produce a large body of scientific, technical, and policy work in targeted core safeguards capabilities and human capital development activities. Not only does the NGSI program carry out activities across multiple disciplines, but also across all U.S. Department of Energy (DOE)/NNSA locations in the United States. However, products are not readily shared among disciplines and across locations, nor are they archived in a comprehensive library. Rather, knowledge of NGSI-produced literature is localized to the researchers, clients, and internal laboratory/facility publication systems such as the Electronic Records and Information Capture Architecture (ERICA) at the Pacific Northwest National Laboratory (PNNL). There is also no incorporated way of analyzing existing NGSI literature to determine whether the larger NGSI program is achieving its core safeguards capabilities and activities. A complete library of NGSI literature could prove beneficial to a cohesive, sustainable, and more economical NGSI program.

The Safeguards Platform for Automated Retrieval of Knowledge (SPARK) has been developed to be a knowledge storage, retrieval, and analysis capability to capture safeguards knowledge to exist beyond the lifespan of NGSI. During the development process, it was necessary to build a SPARK training dataset (a corpus of documents) for initial entry into the system and for demonstration purposes. We manipulated these data to gain new information about the breadth of NGSI publications, and they evaluated the science-policy interface at PNNL as a practical demonstration of SPARK's intended analysis capability. The analysis demonstration sought to answer the question, "Who leads research and development at PNNL, scientists or policy researchers?"

The analysis was inconclusive as to whether policy researchers or scientists are primary drivers for research at PNNL. However, the dataset development and analysis activity did demonstrate the utility and usability of the SPARK dataset. After the initiation of the NGSI program there is a clear increase in the number of publications of safeguards products. Employing the natural language analysis tool IN-SPIRE™ showed the presence of vocation- and topic-specific vernacular within NGSI sub-topics. The methodology developed to define the scope of the dataset was useful in describing safeguards applications, but may be applicable for research on other topics beyond safeguards. The analysis emphasized the need for an expanded dataset to fully understand the scope of safeguards publications and research both nationally and internationally.

As the SPARK dataset grows to include publications outside PNNL, topics crosscutting disciplines and DOE/NNSA locations should become more apparent. NGSI was established in 2008 to cultivate the next generation of safeguards professionals and support the development of core safeguards capabilities (NNSA 2012). Now a robust system to preserve and share institutional memory such as SPARK is needed to inspire and equip the next generation of safeguards experts, technologies, and policies.

Acronyms and Abbreviations

DOE	U.S. Department of Energy
ERICA	Electronic Records and Information Capture Architecture
FoIS	Foundations of International Safeguards
FY	fiscal year
IAEA	International Atomic Energy Agency
NGSI	Next Generation Safeguards Initiative
NNSA	National Nuclear Security Administration
OSTI	Office of Scientific and Technical Information
PNNL	Pacific Northwest National Laboratory
R&D	research and development
SPARK	Safeguards Platform for Automated Retrieval of Knowledge

Contents

Figures

Tables

1.0 Introduction

In its first five years, the National Nuclear Security Administration's (NNSA) Next Generation Safeguards Initiative (NGSI) sponsored more than 400 undergraduate, graduate, and post-doctoral students in internships and research positions (Wyse 2012). In the past seven years, the NGSI program has, and continues to produce a large body of scientific, technical, and policy work in targeted core safeguards capabilities and human capital development activities. Not only does the NGSI program carry out activities across multiple disciplines, but also across all U.S. Department of Energy (DOE)/NNSA locations spanning the United States. However, products are not readily shared among disciplines and across locations, or archived in a comprehensive library. Rather, knowledge of NGSI-produced literature is localized to the researchers, clients, and internal laboratory/facility publication systems such as the Electronic Records and Information Capture Architecture (ERICA) at the Pacific Northwest National Laboratory (PNNL). There is also no incorporated way of analyzing existing NGSI literature to determine whether the larger NGSI program is achieving its core safeguards capabilities and activities. As the body of nuclear nonproliferation and safeguards knowledge continues to mature, convenient access to these products is necessary to prevent duplication and increase the impact of current and future research. A complete library of NGSI literature could prove beneficial to a cohesive, sustainable, and more economical NGSI program.

NGSI products and other safeguards-related resources are available from a variety of sources, including peer-reviewed journals, conference proceedings, video aggregators, or through professional connections across organizations. Some efforts have been made to compile these resources into repositories; however, those efforts have been either broadly focused without specificity for safeguards or narrowly focused for a single program area and are not designed to be used communally. Efforts include the Office of Scientific and Technical Information (OSTI) and Foundations of International Safeguards (FoIS) project. There are also access-controlled resources such as the NA-24 eRoom that provide additional access to non-published works and distribution-restricted information (e.g., Official Use Only). Since there is no general repository for such literature, current accessibility of these resources relies upon the ability of a researcher to parse a large amount of publications unrelated to safeguards. Furthermore, an understanding of the breadth of safeguards-related research is largely localized to those researchers working closely with a project (i.e., collaborators or editors) and those who have read published literature (often limited to those who worked on the project or were heavily involved with it). Without an aggregated library, the research outputs of the NGSI program and the larger safeguards community are often lost in a sea of publications.

The Safeguards Platform for Automated Retrieval of Knowledge (SPARK) has been developed to be a knowledge storage, retrieval, and analysis capability to capture safeguards knowledge to exist beyond the lifespan of NGSI. SPARK will serve as a central, controlled location of contemporary work products and training materials being produced by the NNSA-supported nonproliferation community and provide a point-of-entry for this community to access and maintain institutional knowledge. Additional capabilities include sentiment analysis, rhetorical analysis, and automated discovery of new information through harvesting. All of these capabilities will serve NGSI's Human Capital Development need of knowledge retention related to program interest areas.

This report discusses an initial SPARK training dataset (a corpus of documents) for initial entry into the system and a demonstration of SPARK as a source of data for analysis. The methodology and scope of documents included in the demonstration are described. The report also demonstrates how to manipulate the data to gain new information about the breadth of NGSI publications. Finally, it empirically evaluates the science-policy interface at PNNL as a practical demonstration of SPARK's intended analysis capability. The analysis demonstration considers the question: "Who leads research and development (R&D) at PNNL, scientists or policy researchers?"

The analysis was inconclusive as to whether policy researchers or scientists are primary drivers for research at PNNL. However, the dataset development and analysis activities did demonstrate the utility and usability of the SPARK dataset. These activities highlighted the importance of the NGSI program in increasing and maintaining regular publication of safeguards products. The analysis emphasized the need for an expanded dataset to fully understand the scope of safeguards publications and research both nationally and internationally. The development of the dataset also accentuated the need for an automated system to streamline sustainable curation. A robust system to preserve and share institutional memory is needed to inspire and prepare the next generation of safeguards experts, technologies, and policies. SPARK could be that mechanism.

2.0 Background

NGSI was established in 2008 to revitalize the international safeguards technology and human resource base by leveraging U.S. technical assets and partnerships to keep pace with demands and emerging safeguards challenges (NNSA 2012). It was charged with developing the policies, concepts, technologies, expertise, and infrastructure necessary to sustain the international safeguards system as its mission evolves over the next 25 years. NGSI promotes an interdisciplinary approach to human capital development and safeguards R&D through five main pillars: Policy and Outreach, Concepts and Approaches, Technology Development, Human Resource Development, and Infrastructure Development (NNSA 2009). Interns and research associates conduct new and continuing research, activities, and technological developments, which result in the preparation of a large number of products including papers, reports, presentations, videos, etc. that span all safeguards-related disciplines. However, there is no broadly available repository of NGSI products, so much of this content is difficult to find among so many publications.

In fiscal year (FY) 2014, the PNNL proposed a project to adapt existing content management tools to create SPARK, which is focused on the challenges specific to the NGSI context. The proposal called for execution of three tasks[1] with the following deliverables:

- Task 1: (Funded and completed in FY 2014.) Developed a scoping-study/strategy document that informs the potential deployment of SPARK across the NGSI program areas. It included a discussion of the full SPARK capability set and identified potential roadblocks with regards to information security (Toomey et al. 2014).
- Task 2: (Executed using NGSI research associate funding, documented by this report.) Develop a curated document library of PNNL resources and training materials for input into the SPARK website, when completed in Task 3.
- Task 3: Develop the complete SPARK website for demonstration with PNNL materials developed in Task 2, which includes a
  - demonstration of upload and tagging capability for document collection/collation
  - demonstration of automated document search and collection capability
  - demonstration of completed, access-controlled, PNNL library
  - document describing contents of training material catalog and identification of gaps in the available training material.

Task 1 was completed in the fourth quarter of FY 2014. The strategy document informs the potential deployment of SPARK across the NGSI program areas. It provided a detailed set of design guidelines and justifications, highlighted critical information security considerations, and detailed the timeline and budget requirements to fulfill the requirements and execute Tasks 3. Outstanding requirements include the development and testing of the full user interface and development of a complete safeguards-product library. The estimated time for completion of the demonstration is two years from the start of Task 3.

[1] For the purposes of this document, tasks refer to three steps defined in the scope of work; phases refer to sub-tasks within a given task.

Deployment to the laboratory complex is expected to take an additional 9 months. An alternative option requiring a smaller development budget was also proposed. Funding has yet to be approved for Task 3.

This report is the deliverable for Task 2, which describes the creation of a SPARK training dataset. For testing purposes, it is initially limited to all safeguards-related products produced at PNNL.

2.1 Introduction to ERICA, the Data, and Dataset

The SPARK training dataset comprises all safeguards-related PNNL-authored products submitted to ERICA, an online system that automates and tracks the information release process and serves as a library of all unclassified PNNL products. It encapsulates products that are slated for release outside the laboratory—whether to a client or the general public. The term "products" is used to describe the written entry into ERICA, regardless of product type. Product type categories include: abstract, abstract/presentation, brochure/flyer, conference paper, exhibit, book chapter, book/conference proceedings, journal article, other, presentation, formal report, and video. Each entry includes descriptive information (e.g., author[s], product type, date, etc.) and a paragraph describing the work (e.g., an executive summary or abstract). The full products are sometimes digitally attached to their ERICA entries, but housed in a separate system for information security reasons. In order to gain full access to a product, the inquirer must submit a request through the PNNL Technical Library, contact the authors directly, or check public sources. For now, the SPARK training dataset will not include the final product. Inclusion of final products would require further research on how to address information security concerns.

The entries in ERICA do have some weaknesses that SPARK will seek to overcome. Since authors submit their own entries and the quality relies on the effort they put forth in completing the ERICA entry, some product types (abstracts, presentations, and reports) are typically more complete than others.[1] Furthermore, authors often submit multiple product types on the same topic (i.e., an abstract, presentation, and report) to maximize the exposure of their project. Thus, approximately 40 to 50 percent of entries are duplicates.[1] Moreover, no product type is all-inclusive; some products are only produced as a technical report, for example, but may not have an abstract. In addition, an entry's publication date describes the date the product was approved for external release; it does not necessarily indicate the status of the project itself. These weaknesses create complications when analyzing entries collectively, because some topics may be over-represented, product descriptions may be incomplete, or dates may not reflect a milestone in a project (when looking at time series). To adjust for this issue, SPARK will allow the user to analyze within (as well as across) each product type.

The SPARK dataset is intended to include products across all years. The training dataset is restricted by the entries in ERICA and the timing of this project. It includes products with publication dates between September 1, 2005 and August 31, 2013. ERICA was introduced in FY 1999, but archives for the first few years are somewhat incomplete.[1] Thus, we limited the products to those after 2005. August 2013 corresponds with the initiation of this project. As we transition to a more permanent dataset, new and archived products will be included as they become digitally available.

[1] Parker, Tomiann. 2014. Interview of Tomiann Parker, (Pacific Northwest National Laboratory) by A Sayre (Pacific Northwest National Laboratory) "Information Release Discussion," 2 July 2014.

2.2 Establishing the Dataset

ERICA contains almost 9,500 entries and includes the full scope of publication topics at PNNL. For SPARK, it was necessary to draw out only the safeguards-related products. It is difficult to draw a clear line among projects that have a safeguards application and those that do not. The first task was to build a model of safeguards terminology and then apply that model to extract only the records with potential connections to safeguards. We then used the natural language processing software IN-SPIRE™ to refine the dataset.

2.2.1 Defining the Dataset Parameters: Safeguards and Safeguards-related

Safeguards, as it relates to nuclear weapons, dates back to the 1950s and the establishment of the International Atomic Energy Agency (IAEA). The purpose of safeguards, as defined by the IAEA, is to "provide credible assurances that States are honoring their safeguards agreements…and to…detect early any misuse of nuclear material or technology" (IAEA 2014b). For this report, "safeguards" are any activity that supports the aforementioned purpose. The general criteria for inclusion of a product in this dataset are that it describes safeguards or a safeguards-related activity, or has an application in safeguards.

To add more detail to the IAEA's definition of safeguards, we developed a word tree, with safeguards as a main branch. The tree breaks safeguards down into categories of research and activities. These were further broken down into more specific sub-categories and keywords, which together would describe a safeguards-related activity. There are several parallel trees which are related, but distinct from safeguards activities (e.g., security and safety). Figure 2.1 shows the safeguards branch of the word tree, which breaks down safeguards into its most fundamental activities. The asterisk (*) denotes that there is an additional tree which further break down that activity. Figure 2.2 is a sub-tree of the safeguards tree, and a continuation of nondestructive assay (NDA). Safeguards also had several parallel trees, which are related, but distinct from safeguards activities (e.g., security).

Expert interviews and open source literature assisted the development of safeguards categories. Within the literature review, the IAEA safeguards glossary, the *International Target Values 2010 For Measurement Uncertainties in Safeguarding Nuclear Materials* (IAEA 2010), and *Nuclear Safeguards, Security, and Nonproliferation* (Doyle 2008) were particularly helpful.

The process of defining safeguards is somewhat imperfect because the definition of safeguards is broad and the activities which could have safeguards applications are also widespread. To avoid unnecessary exclusion, we use a broad definition of safeguards. The definition of safeguards and keyword groupings were refined with each coded product throughout the categorization process. This process will be discussed in greater detail in Section 2.2.3.

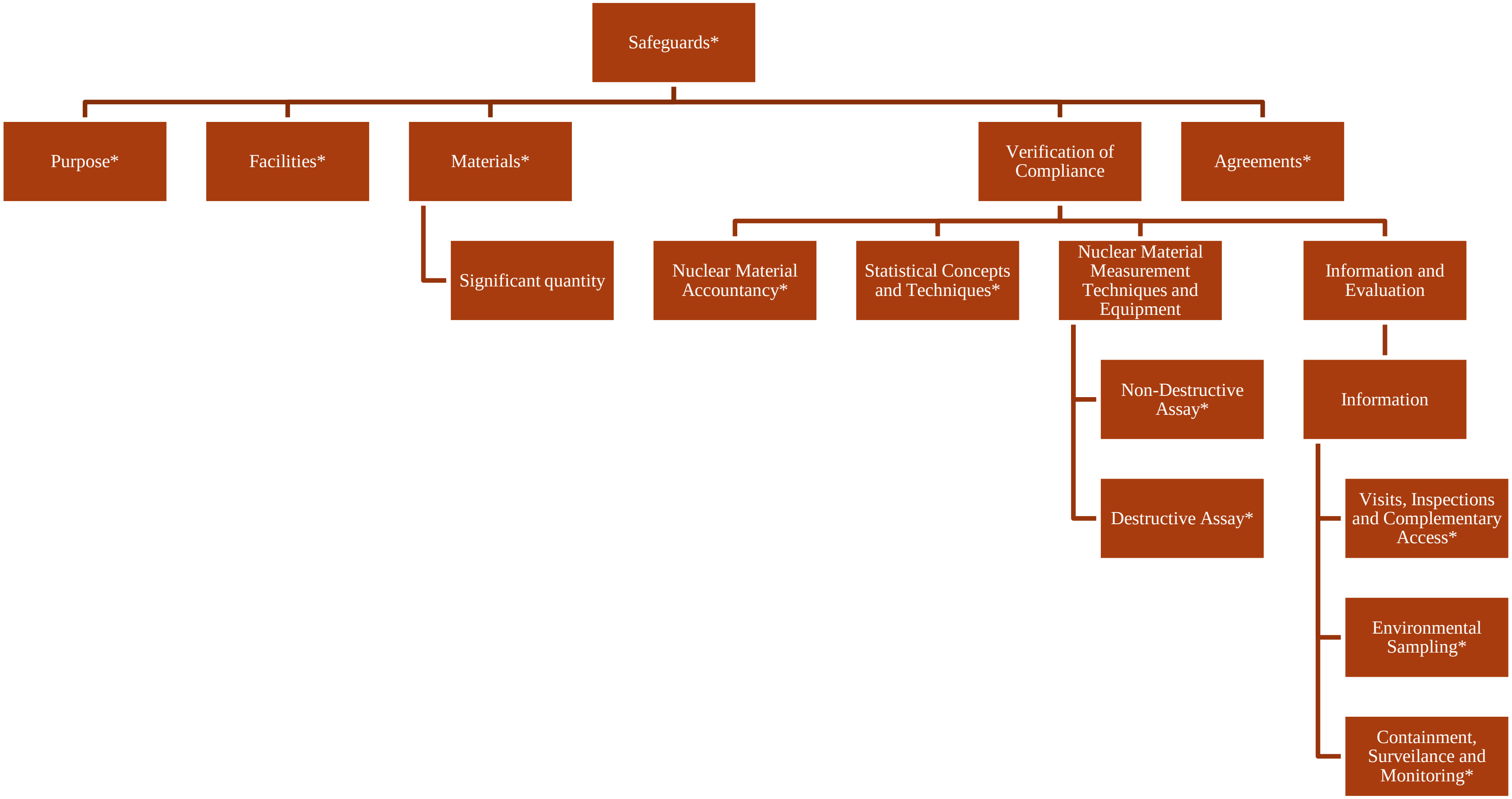

Figure 2.1. Safeguards Word Tree

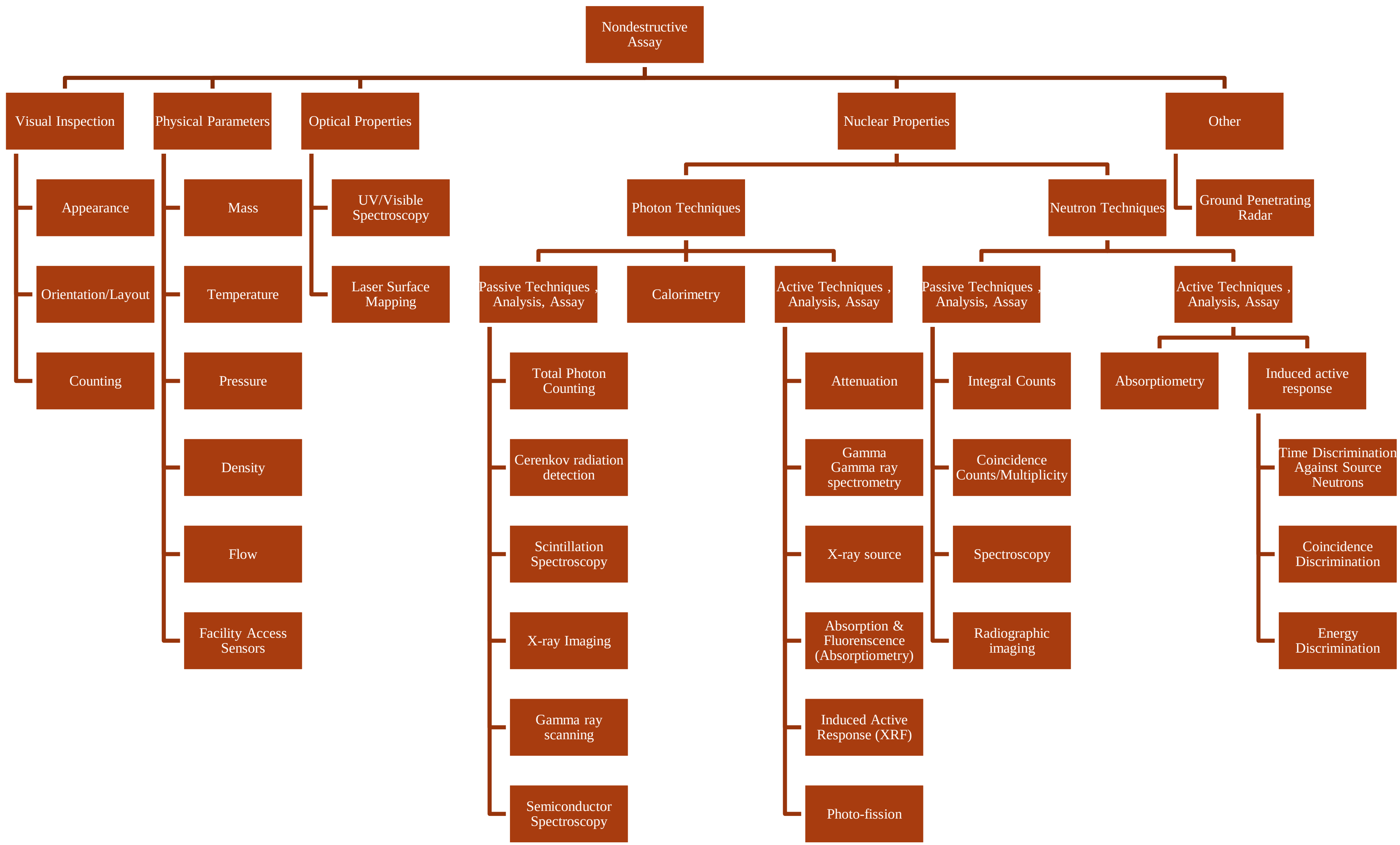

Figure 2.2. Nondestructive Assay Sub-Tree

2.2.2 Tools for Establishing the Dataset

2.2.2.1 Python

Python is a general-purpose, high-level programming language, which was used to extract products from the ERICA Database using a list of keywords. The keyword list was developed using expert interviews and general safeguards and nonproliferation knowledge. It contains the most general terms for safeguards and related activities. The keyword list is included in Table 2.1. Due to missing data and the possibility that authors would not include appropriate keywords, the script pulled all products by authors who have produced at least one product containing a designated keyword. Instead, this method assumes that PNNL authors who write about safeguards have written at least one paper, which contains a keyword. The Python script yielded 8,556 entries. At this point, the dataset still contained entries far beyond the scope of this project (i.e., those without safeguards applications) and it required further distillation.

Table 2.1. Keywords Used in Python Script

Keywords	Keywords
nuclear, safeguard[(a)]	nuclear, noncompliance
prolif[(a)], nuclear	nuclear, security
global regime	fuel cycle
confidence building measure[(a)]	nuclear, warhead
cooperative border security program	fissile material
domestic counterproliferation enforcement	WMD
global initiative[(a)] for proliferation	weapon[(a)] of mass destruction
prevention	nuclear
export control	atomic
Additional Protocol	

[(a)]Denotes that that different variations of the word are allowed (e.g., for "tag[(a)]", tag and tags are permissible)

2.2.2.2 IN-SPIRE™

IN-SPIRE™ was the main tool for understanding the products, focusing the dataset, and categorizing the products. It is an information visualization software tool developed and distributed by PNNL for use by Battelle/PNNL staff members, the U.S. federal government, and commercial customers. IN-SPIRE uses word proximity and composition to synthesize large sets of unformatted text documents. It clusters similar documents and provides keywords and themes, which can reveal relationships between documents. It creates themes of multiple keywords that appear in the same or similar context between documents; this function allows a researcher to see how documents are grouped. Users can also create new groups of documents. The text document can be delimited or divided into unique sections (e.g., author, title, summary, etc.) for analysis. Other IN-SPIRE tools help to explore trends over time and relationships between concepts, which we used for this project.

We applied the safeguards definition and word tree developed from the literature and experts to the dataset in IN-SPIRE to parse out the safeguards-related products from the 8,556 that Python extracted from ERICA.

2.2.3 Establishing the SPARK Dataset

We began the product categorization based on keywords (inclusion or exclusion) and then manually categorized ambiguous products. A safeguards activity can usually be defined by a few keywords or phrases, but is highly context-dependent. Table 2.2 gives example keywords for inclusion and exclusion, which were used to categorize large clusters of products in IN-SPIRE. Products that contain keywords from Table 2.1 or the word tree (Figures 2.1 and 2.2) would most likely have safeguards applications and are easy to code. We used a combination of keyword searches and IN-SPIRE's clustering tool to classify large groups of linguistically similar products. Figure 2.3 demonstrates IN-SPIRE's clustering tool. Different combinations of the keywords for inclusion yielded products for the dataset.

Table 2.2. Example Keywords Related to Safeguards and Related Activities

Keywords for Inclusion	Keywords for Exclusion
nuclear energy	climate change[(b)]
export control	technetium
IAEA	migration
safeguard[(a)]	nuclear waste management[(c)]
prolif[(a)]	migration and nuclear waste
nondestructive assay (NDA)	"nuclear energy" and contaminant and environment
destructive assay (DA)	geological repositories
"MC&A" or "material control and accounting"	environmental remediation
Additional Protocol (AP)	environmental contamination
SSAC/RSAC – "state system"	radioactive waste management
HEU or LEU	vadose zone
trafficking	nuclear fuel management; irradiated fuel management; nuclear fuel recycling
arms control	nuclear fuel reprocessing; waste form
enriched, enrichment	geologic[(a)] repositor[(a)] and corrosion
border security	CTBT[(d)]
tamper indicat[(a)], "tag[(a)]" and "seal[(a)]"	
nuclear forensics	

[(a)]Denotes that that different variations of the word are allowed (e.g. for "tag[(a)]", tag and tags are permissible).
[(b)]Although climate change has implications for the nuclear energy industry, it only has a tenuous relationship with safeguards. Eventually, research may find that climate change has an effect on safeguards implementation. At this time, current climate change research has explored the possible effect to national security, nonproliferation, and nuclear energy, but there are no direct links to safeguards.
[(c)]The effects of this waste on the environment are unrelated to safeguards topics. Conversely, the detection of radioactive material in the environment has applications for detecting clandestine activities (included).
[(d)]The Comprehensive Test Ban Treaty is not related to safeguards. The technology regarding detection of nuclear tests is related to technologies used to detect undeclared activities. Much of the detection and international monitoring station literature was included.

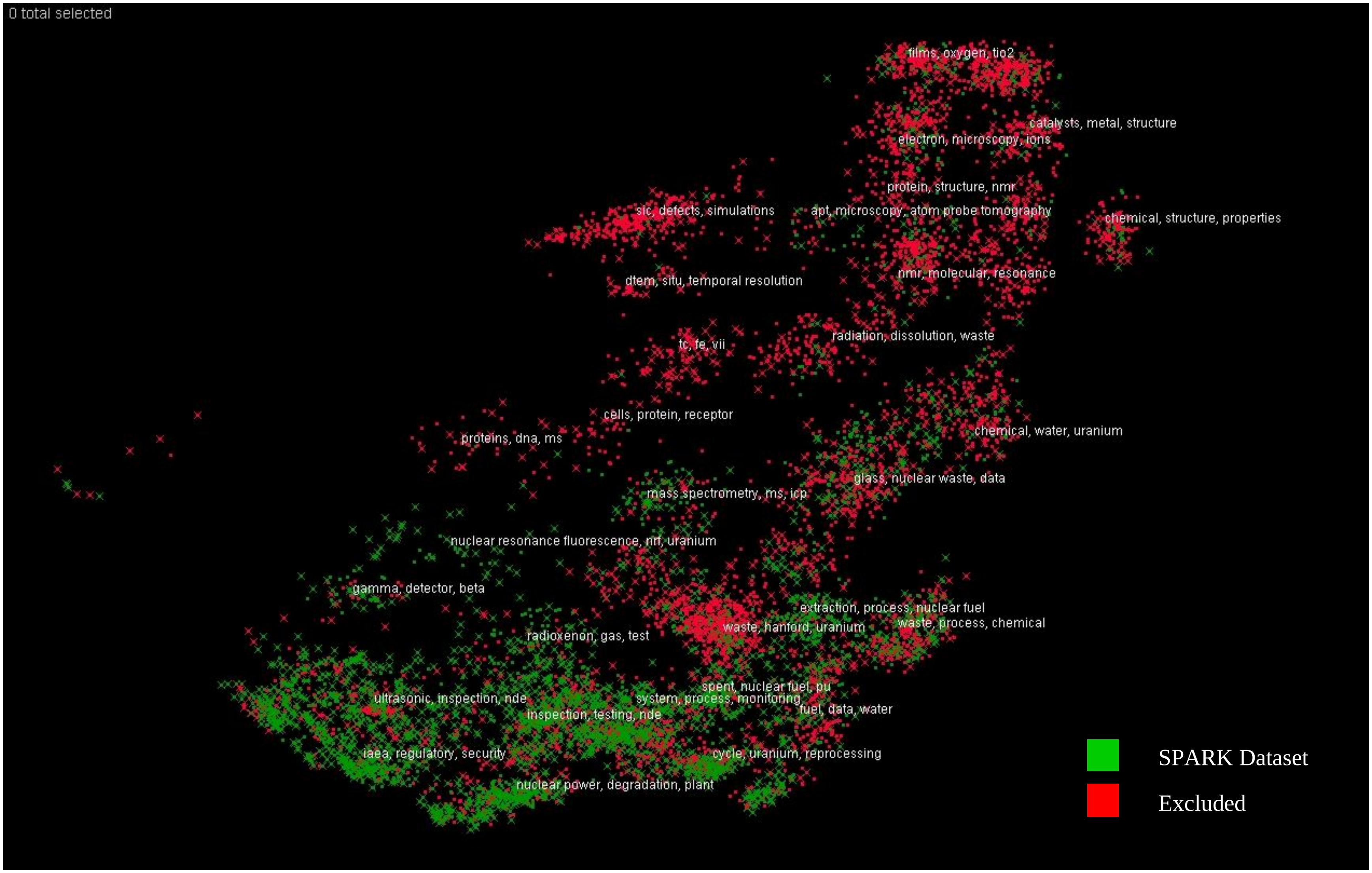

Figure 2.3. IN-SPIRE Product Clusters

For instance, products containing both “safeguards” and “IAEA” and those clustered nearby were included. A safeguards application can also be ruled out based on critical keyword groupings. The fact that a product contains one of the excluded keywords does not necessarily mean that it is irrelevant. A product must generally describe one of the excluded activities and not contain safeguards-related keywords for it to be excluded. Keywords successfully categorized 5,295 products.

Many products describe, but do not name one of the activities in Table 2.1 or the word trees. Those documents require critical science and technology knowledge to code. They also required a manual approach with more general search terms. One method involved incremental refinement of search terms. By focusing search terms from the most general to the most specific, classifying whole groups of products became much easier. For example, after searching for a general term like “verification,” a researcher can search within those products for an additional keyword, like “diversion” to obtain only the most relevant documents. This method helped accurately categorized those products whose safeguards applications were unclear because of word usage. For example, “safeguards” could be used in a more traditional context, meaning “to protect,” rather than in reference to IAEA safeguards. The manual categorization also made use of IN-SPIRE’s cluster feature. By first isolating a cluster and reading a few products in that cluster, keywords were identified and the entire group was categorized accordingly. The manual process was iterative, but necessary to classify 3,261 products.

Figure 2.4 is a decision tree that describes the designation methodology, which involved software decision making supported by manual categorization. This effort demonstrated the need for an automated parsing system, which the training dataset will be used to build. A large amount of time was devoted to the background research and detail checking required to resolve manually entered keywords by disparate sets of researchers, showing that any human-managed system will lead to critical data management problems.

Once the documents were coded into safeguards and non-safeguards categories, the resulting SPARK dataset contained 3,750 products out of the original 8,586 provided by Python. Figure 2.5 shows the products categorized by each tool and the final SPARK dataset.

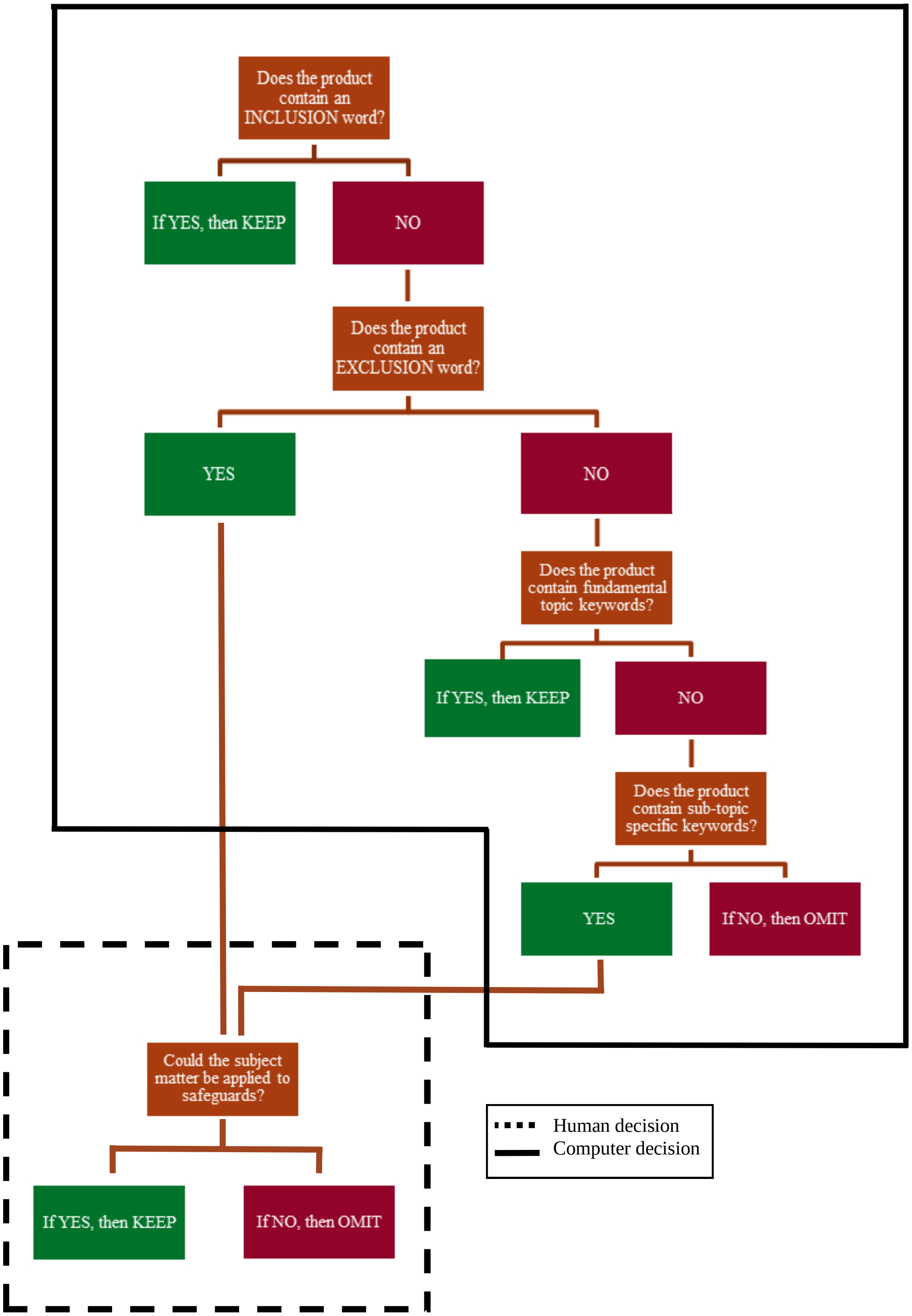

Figure 2.4. Categorization Decision Tree

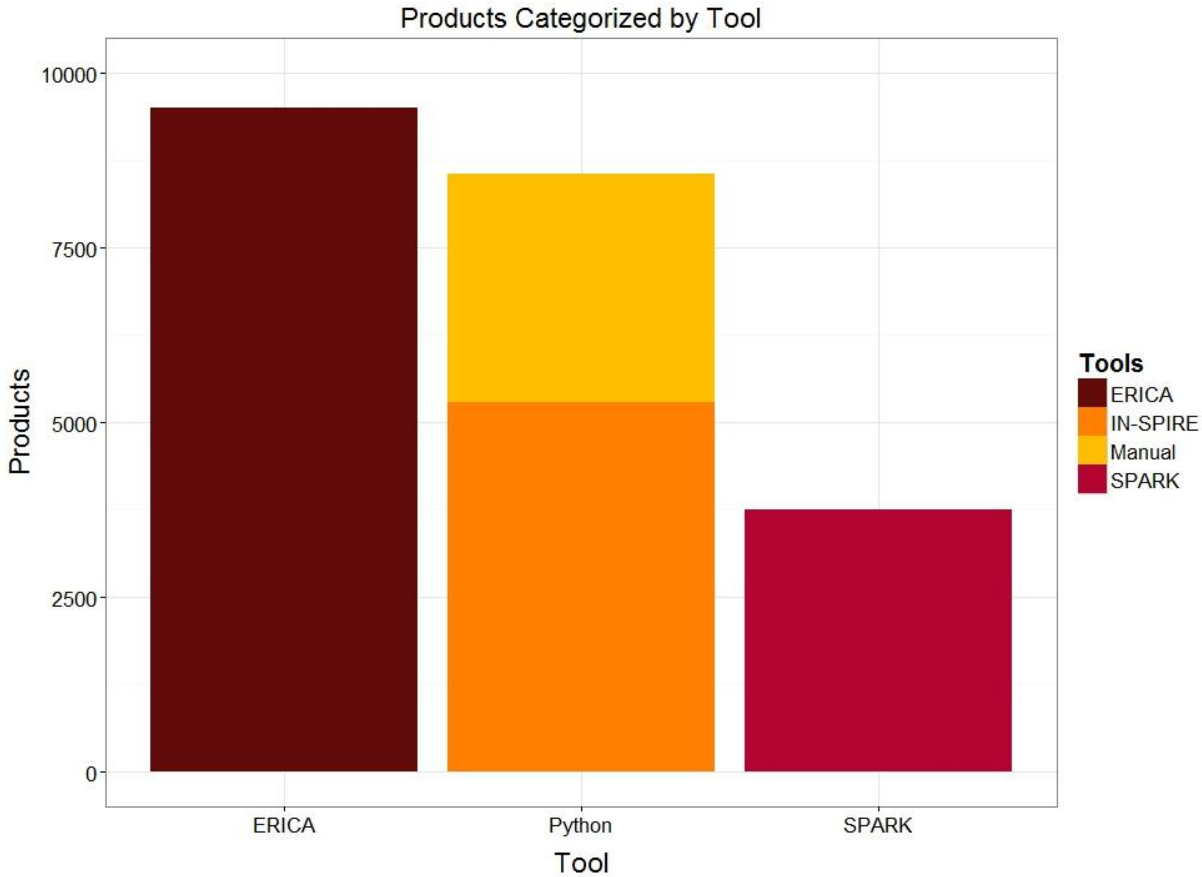

Figure 2.5. Products Categorized by Tool

2.3 SPARK Dataset

Table 2.3 and Figure 2.6 provide more information on the distribution of products in the SPARK dataset. Presentations, abstracts, conference papers, technical reports, and journal articles (in that order) make up a substantial portion of the products.

Table 2.3. SPARK Dataset Summary Statistics

	Count (% of total)
Total Products	3750
Abstracts	1012 (27%)
Presentations	1401 (37%)
Reports	369 (10%)
Conference Papers	417 (11%)

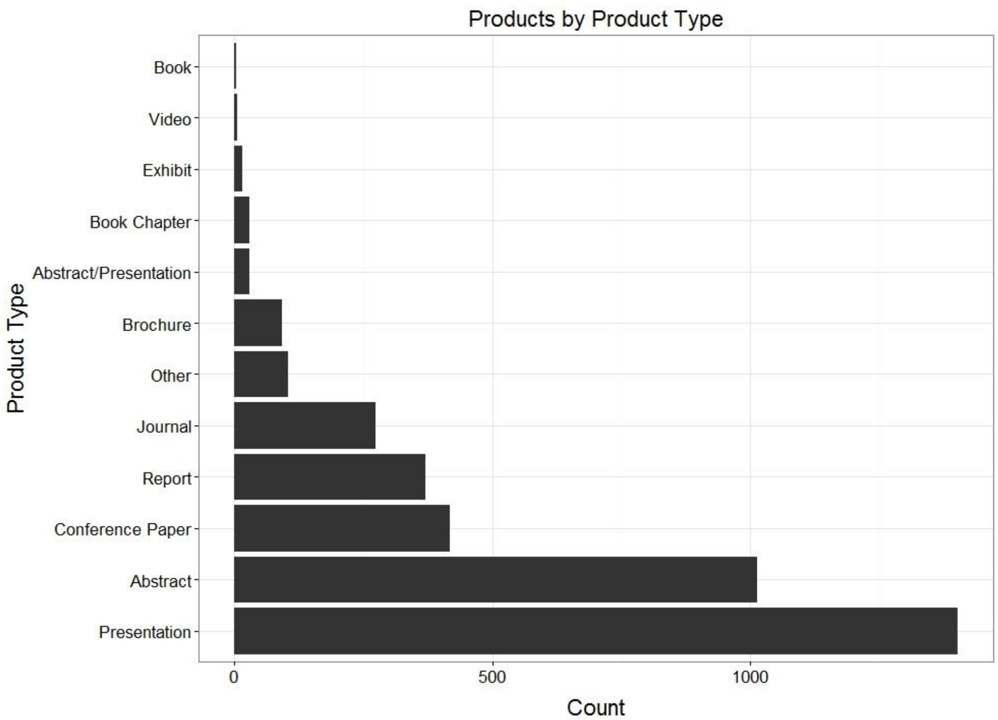

Figure 2.6. Products by Product Type

Figure 2.7 shows the distribution of the SPARK dataset over time. Product submission to ERICA is clearly seasonally and monthly driven. December has the fewest submissions, while the spring and summer months have the most. They also peak around months with the due dates for conferences. The sharp increases in February, June, and July correspond with the submission requirements for the Institute of Nuclear Materials Management's (INMM) Annual Meeting. INMM has due dates in February (for abstracts), June (for conference papers), and July (for presentations/posters). The distributions per product are presented in Figures 2.8, 2.9, and 2.10. The American Nuclear Society (ANS) also has their abstract deadline in June, and PNNL has a safeguards course as well at this time. The relationship between government participation in these conferences and publication rates might be worth researching further.

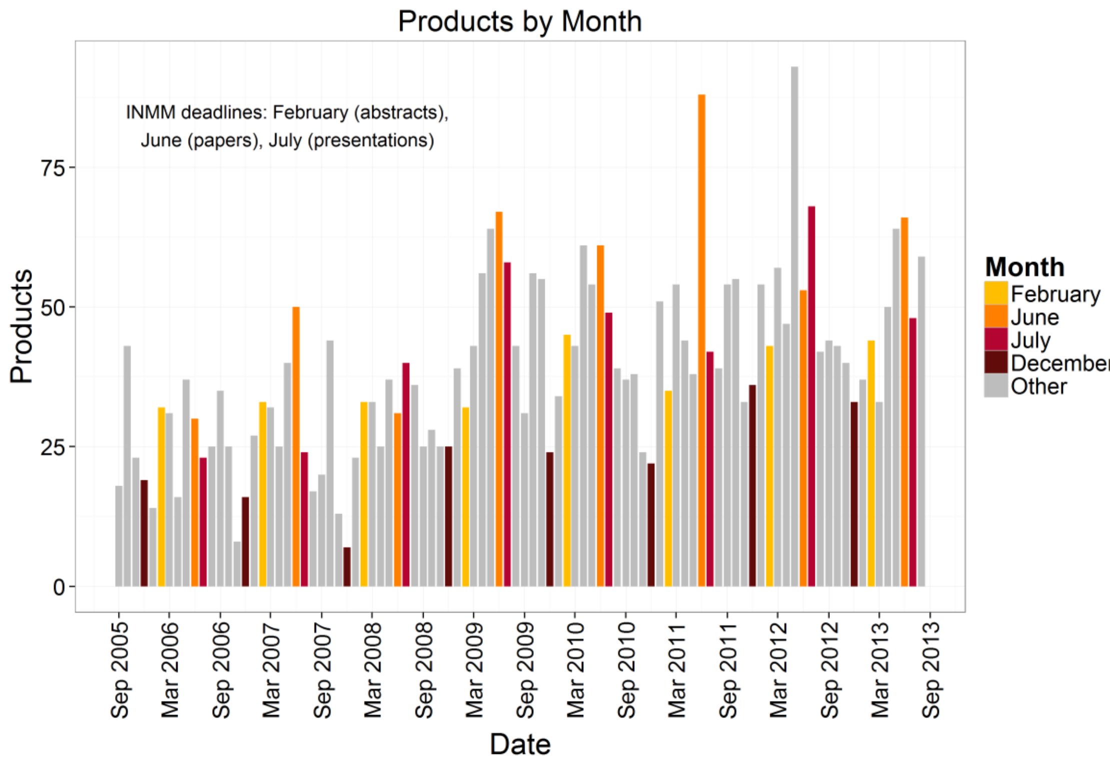

Figure 2.7. All Products Submitted to ERICA by Month

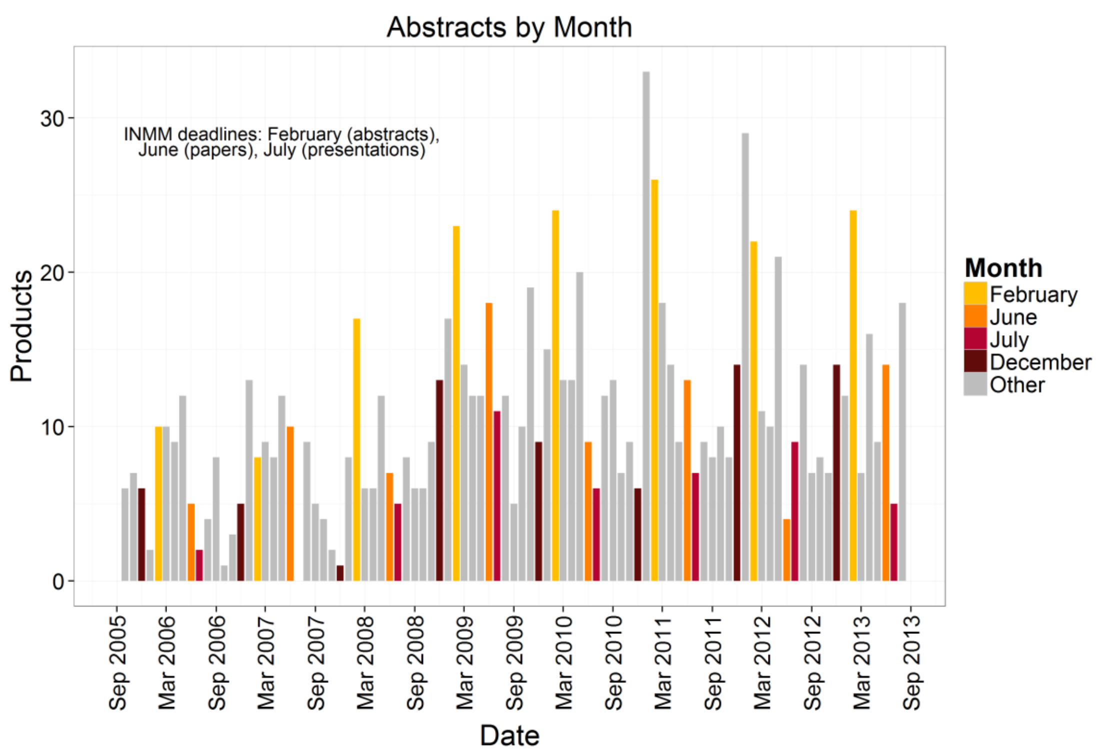

Figure 2.8. Abstracts Submitted to ERICA by Month

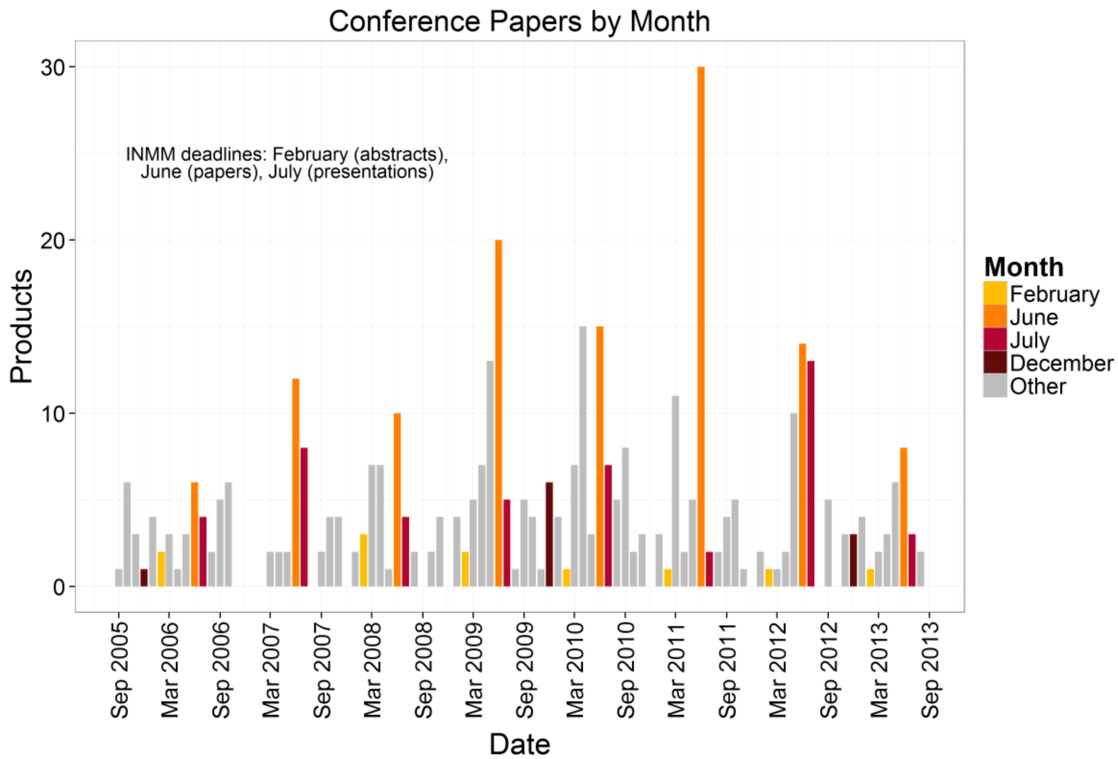

Figure 2.9. Conference Papers Submitted to ERICA by Month

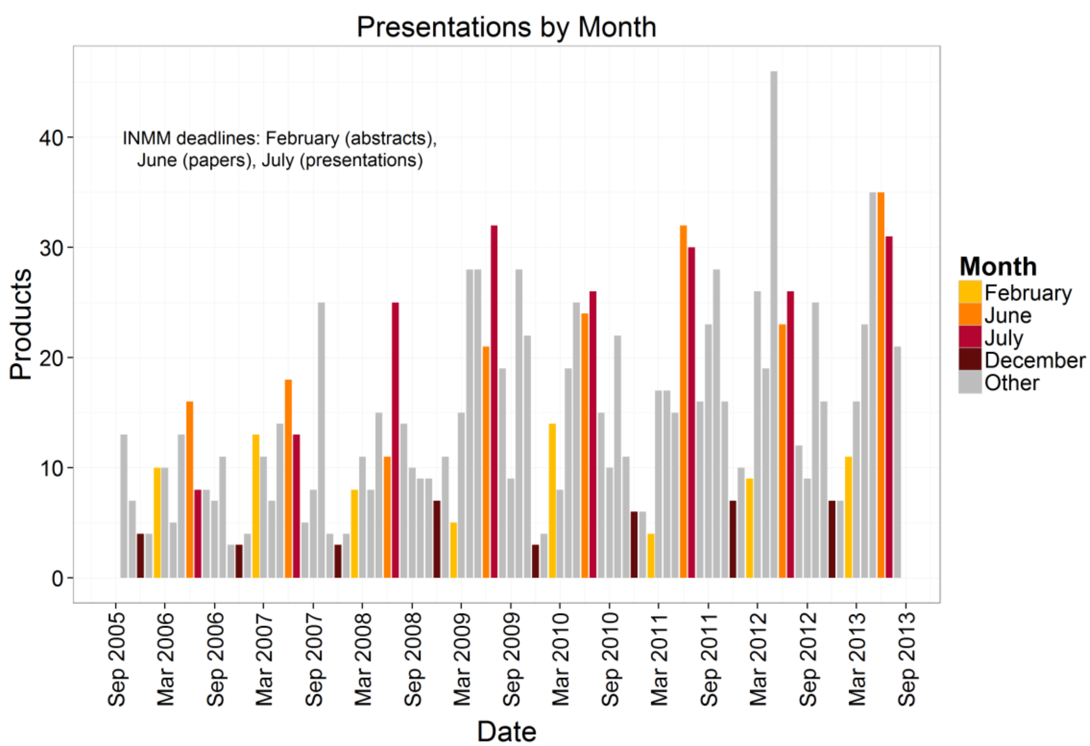

Figure 2.10. Presentations Submitted to ERICA by Month

2.3.1 SPARK Research and Analysis Possibilities

2.3.1.1 Coding the Dataset: Policy and Technical Product Designation

SPARK dataset is intended to enable researchers to discover and share safeguards literature within the community. An additional use is to aggregate the information for analysis to understand trends in safeguards research. To do that, researchers may need to further code and subset the data. Coding the data allows the researcher to apply attributes to each data point to facilitate analysis. For this example, we demonstrate how to designate the products as policy or technical products in the SPARK dataset.

The methodology for designating products as either policy or technical is similar to that for the SPARK dataset, but with different criteria. It was a process of exclusion and inclusion using IN-SPIRE's clustering tool to classify large groups of linguistically similar products followed by manual categorization. Partway through, it became clear that some products were neither policy nor technical in nature, but provided general information. We created a new category called knowledge/training (knowledge) to reflect this finding. Policy and technical products provide something new to the community, while knowledge products provide descriptions of existing research. The following definitions explain the criteria for classification in respective categories:

Policy. A policy product provides new analyses to the political or technical communities. These products either evaluate a given policy or technology in relation to given political or institutional environment or international relations. Policy literature describes regulation or procedure and is typically qualitative in nature. Policy products can evaluate a given technology or technologies for their political implications.

Technical. A technical or scientific product presents or tests a physical or biological science property or technique. Technical literature also includes data analysis and social modeling tools. The research methods are quantitative and based on the scientific method. Technical papers can also evaluate a technology's technical effectiveness for a given purpose.

Knowledge/Training. This category encompasses general literature, which does not add anything new to a particular discussion. Overviews and summaries of safeguards-related activities would be incorporated in this category. Most of this literature has been presented to educate rather than supply analyses or draw new conclusions. Some examples include an overview of safeguards and the IAEA, a summary of nondestructive assay activities at PNNL, and the applications of detection capabilities, where the capability is not new.

Table 2.4 provides information on the breakdown of the SPARK dataset by policy, technical and knowledge/training products.

Table 2.4. SPARK Dataset Summary Statistics for Policy/Technical Coding

	Count (% of total)
Total Products	3750
Technical	2530 (67%)
Policy	732 (20%)
Knowledge/Training	488 (13%)

Figure 2.11 presents the distribution of products by designated category and product type, while Figure 2.12 shows the distribution by category and fiscal year. Technology products make up the majority, followed by policy and knowledge. Policy and technical products are distributed much along the same lines by product type, but knowledge has a much higher number of brochures relative to the other two.

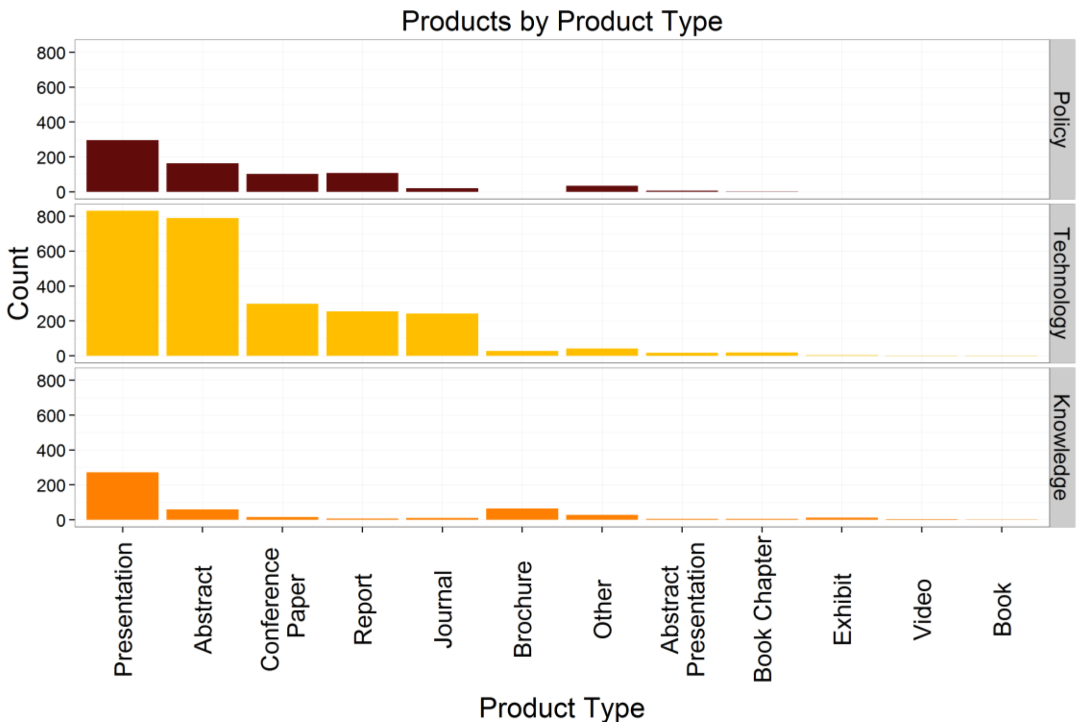

Figure 2.11. Products by Product Type

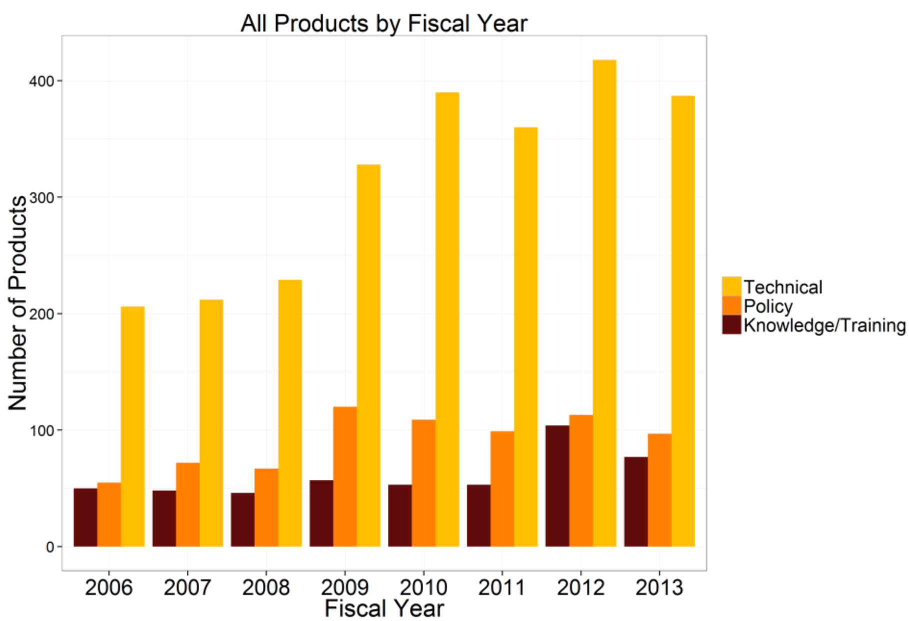

Figure 2.12. Products by Fiscal Year and Designation

Figure 2.13 shows the clustering of topics by designation in IN-SPIRE. The clusters demonstrate the presence of vocation-specific terminology. Policy and technical products are not interspersed when visualized in IN-SPIRE. Instead, they are grouped by similar vocabulary. This may seem obvious given IN-SPIRE's methodology. However, it is apparent that policy and technical products have their own terminology. Policy products, regardless of topic, are grouped more closely together, while the technical papers are grouped by theme. This dispersion is likely because of the product's intended audience. Policy products are directed towards a policy-oriented audience. Technical products are directed towards other scientists and engineers. A fully integrated language may cause more difficulties. Policymakers could be overwhelmed by technical language. Technical experts might view a product as unscientific if the language is too informal. However, some integration may be useful. We will look at a subset of the SPARK dataset to find out.

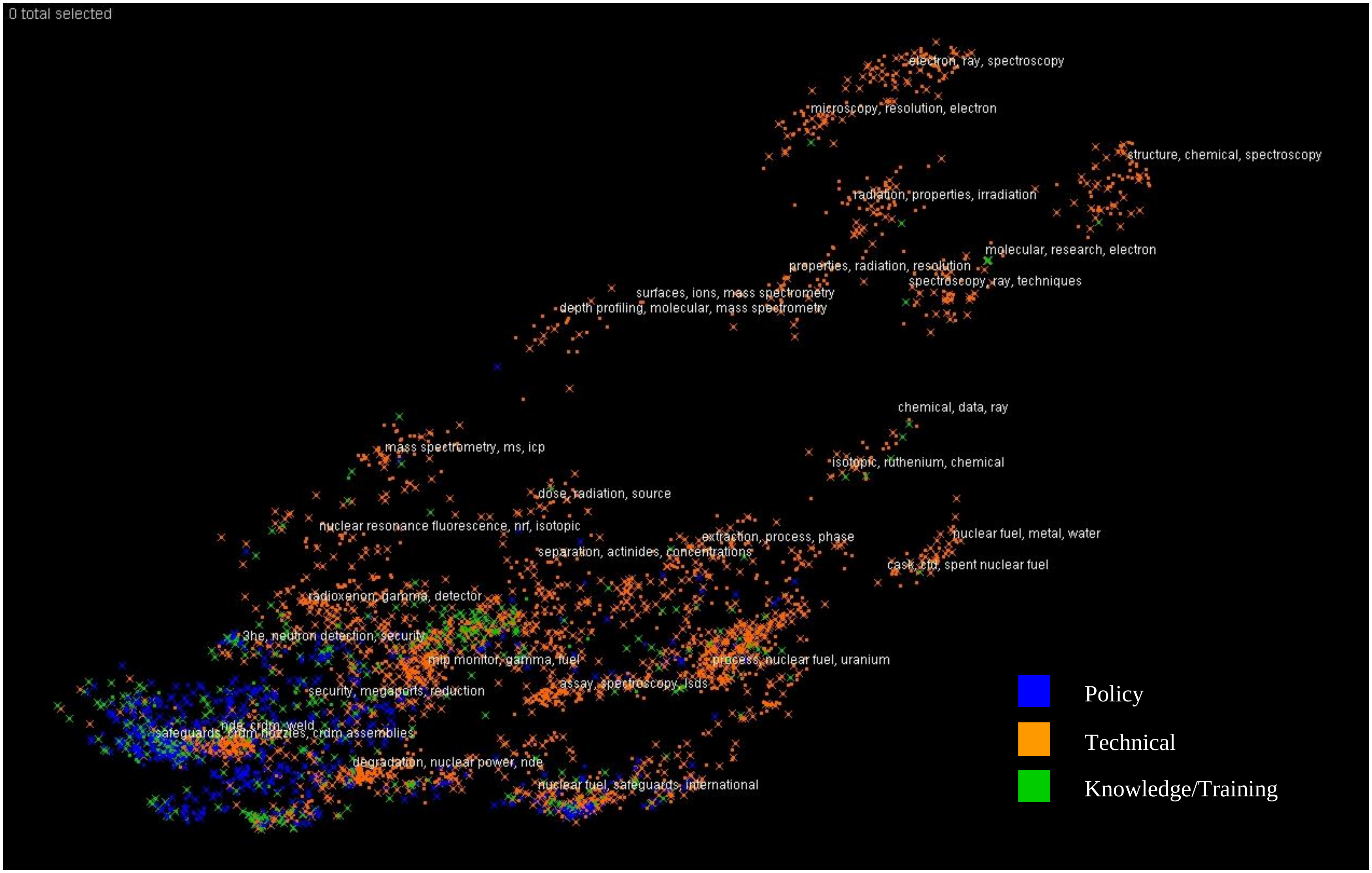

Figure 2.13. SPARK Dataset Coded as Policy, Technical, and Knowledge

2.3.1.2 Creating Subsets: "Safeguards" Subset Designation Methodology

We created a subset of the data to test the sensitivity of results to our coding methodology and to look at word usage of the term "safeguards." This example subset comprises all documents that contain any variation of the word "safeguards" in the context of "international safeguards" and not meaning general "protection." We achieved this by ensuring other safeguards-related terms were also present in the document along with any variation of "safeguards." This also ensured that products did not exclusively use "safeguards" as part of the phrase "Next Generation Safeguards Initiative." Table 2.5 lists these safeguards key terms; the variations of "safeguards" are footnoted. Table 2.6 provides statistics on the subset.

Table 2.5. Safeguards Keywords

Keywords	Keywords
safeguards[(a)]	nuclear material safeguards
safeguard[(a)]	IAEA
safeguardability[(a)]	International Atomic Energy Agency
safeguarding[(a)]	regulate/regulator
safeguarded[(a)]	safeguards perspective
IAEA safeguards	nuclear fuel cycle safeguards
international safeguards	safeguards and arms control
domestic safeguards	education
Next Generation(s) Safeguards Initiative	safeguards and security
safeguards-by-design	safeguards and safety
nuclear safeguard(s)	safeguards challenges
verification	safeguards capacity building
compliance	safeguards-relevant
nondestructive assay	diversion
Additional Protocol	mass spectrometry
safeguards culture	safeguards approaches
accountancy or accounting	nonproliferation and safeguards
safeguards and proliferation	

[(a)]denotes variation of "safeguards"

Table 2.6. Safeguards Subset Summary Statistics

Product Type	Count (% of total)	Category	Count (% of total)
Abstracts	163 (26%)	Technical	330 (52%)
Presentations	215 (34%)	Policy	242 (38%)
Reports	77 (12%)	Knowledge	64 (10%)
Total Products	**636**	**Total Products**	**636**

2.3.1.3 Comparing the SPARK Dataset and Safeguards Subset

The SPARK Dataset and the Safeguards Subset are statistically different in terms of the ratio of both technical-policy-knowledge products and of abstracts-presentations-reports.[1] The safeguards subset has less technical and more policy products than the SPARK dataset. It also has more reports and less abstracts and presentations.

After the NGSI program was announced in September 2008, there is a distinct increase in the number of products in both datasets between fiscal years (FY) 2008 and 2009 (see Figure 2.14). This is even more evident in the subset (see Figure 2.15). International safeguards are a political construction established by the IAEA statute. Therefore, policy products would probably use safeguards as a key phrase more often than technical products, at least in the beginning (2006-2007). Between 2007 and 2008, the number of technical papers, which included the term safeguards increased measurably.[2] This could potentially indicate the integration of policy terminology into technical research. The number of policy products, containing safeguards also increased, but to a lesser extent. In fact, the share of technical products, which used the term safeguards also increased despite the increase in products per year (see Table 2.7). This could be indicative of terminology integration between the policy and technical fields.

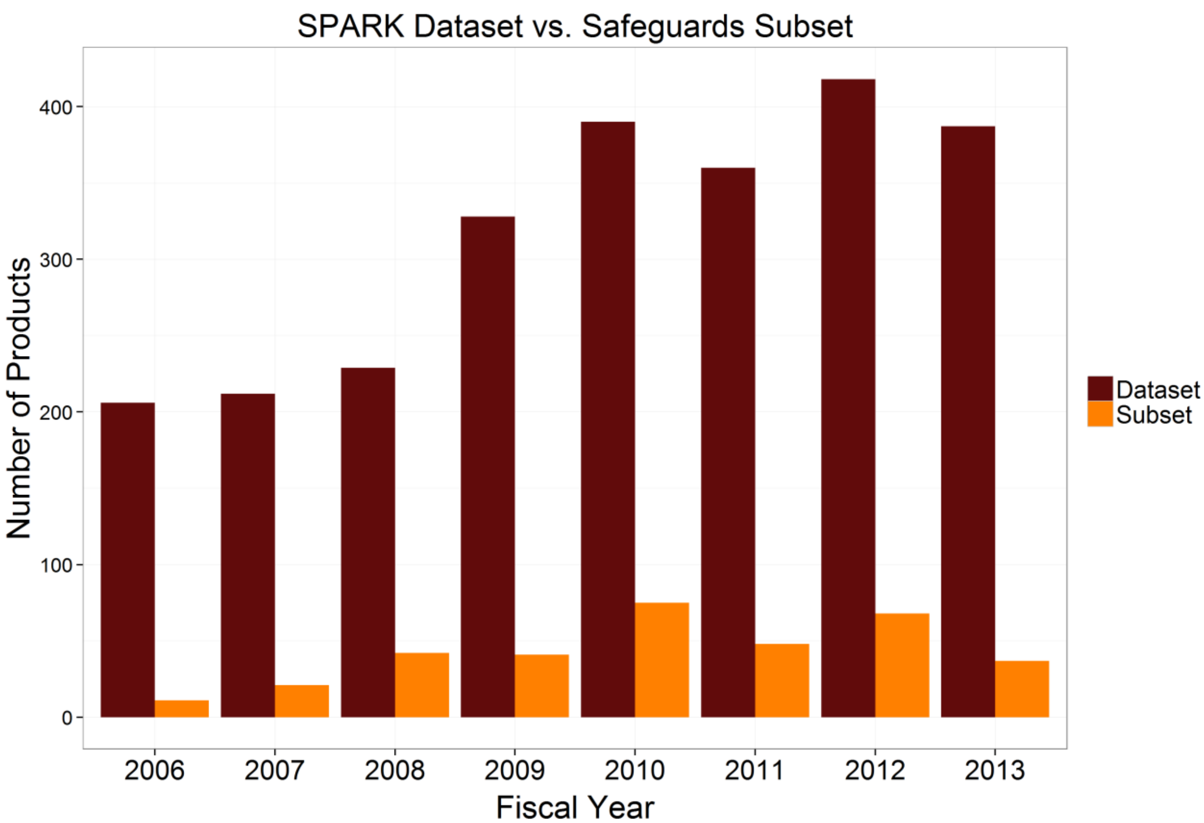

Figure 2.14. SPARK Dataset versus Safeguards Subset

[1] Chi-square tests were significant (p=0.00 and p=0.00 respectively) for product category and product type.
[2] Chi-square test for the expected vs. observed for technical products between 2007 and 2008 were significant (p=0.00).

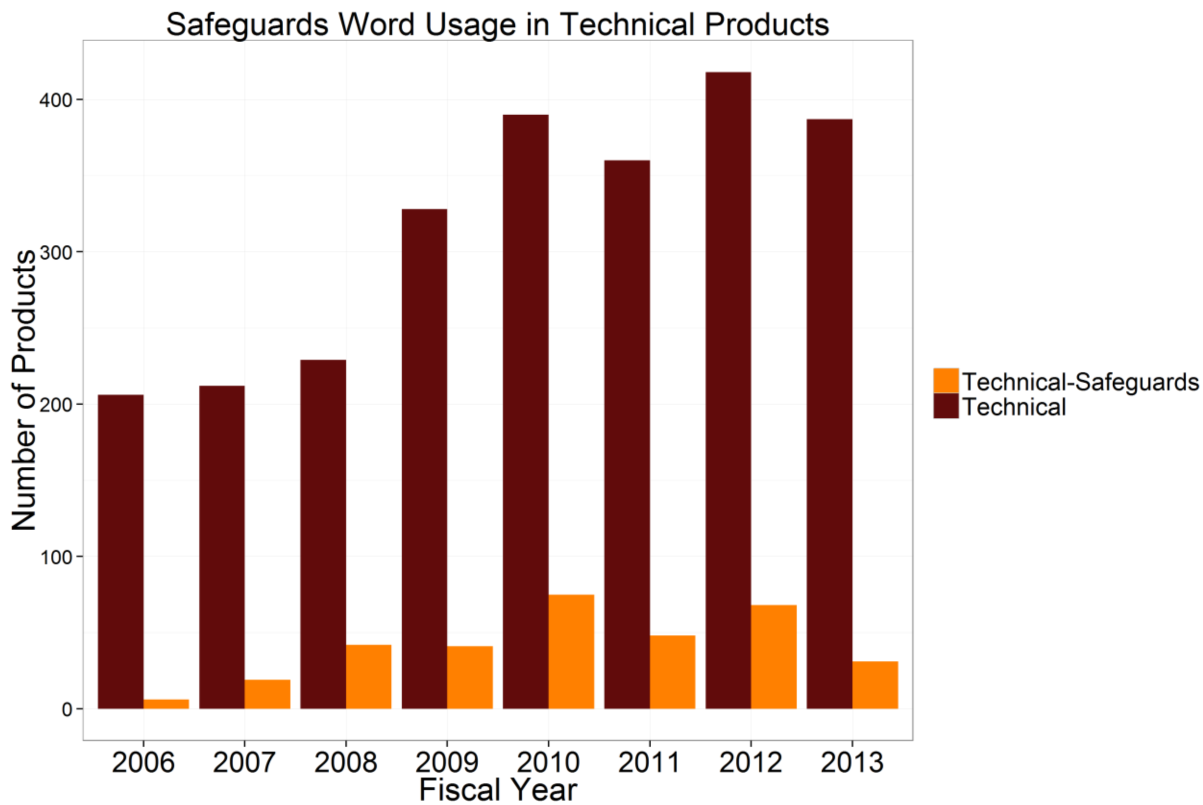

Figure 2.15. Safeguards Word Usage in Policy and Technical Products

Table 2.7. Safeguards Term Usage in Products by Year

		2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Percentage of products that include *safeguards*	Technical Products	3%[(a)]	9%	18%	13%	19%	13%	16%	8%
	All Products	6%[(b)]	14%	18%	19%	24%	18%	18%	13%

[(a)]Read as "3% of technical products include the word *safeguards* (out of all technical products)."
[(b)]Read as "6% of all products include the word *safeguards* (out of all products)."

There are many events and factors, which could have contributed to the decrease in safeguards word usage between 2010 and 2013. Several factors could include travel restrictions because of austerity policies, but more research would be needed to determine leading factors. The SPARK dataset could help researchers understand this type of relationship when combined with other (e.g., budget) information. For now, we will demonstrate a sample application of the SPARK dataset using our coded data and safeguards subset.

3.0 Example Analysis: Drivers of Research at PNNL

Long-time PNNL employees across the policy and technical communities agree that internal driving forces of safeguards innovation include technical/scientific innovation and policy research. Yet, they disagree on which is the primary driver. Scientists generally insist that technical innovation is driving policy research at PNNL, while policy researchers generally claim the opposite. The purpose of this analysis is to understand how safeguards policy work influences the technical work agenda and vice versa. It will also demonstrate a practical application of the SPARK dataset.

3.1 Literature Review

The literature review looks at peer-reviewed articles on agenda-setting research, followed by interviews with personnel on how safeguards-related entities (the IAEA, U.S. government, and PNNL) set their research agendas. Lastly, this literature review explores how individual employees at PNNL might determine their research agendas, because safeguards-specific literature on the topic is absent.

According to Merriam-Webster's online dictionary (2014), an "agenda" is "a list or outline of things to be considered or done." Therefore, a research agenda is an outline of the intended research priorities. Agenda setting is the process by which a list of things (i.e., problems and alternative solutions) is decided (Birkland 2011). In the policy realm, agenda setting is undertaken through debate over priorities (De Boer and Velthuijsen 2001). The group that successfully defines the problem will ultimately define the solution (Schattschneider 1975). In the basic scientific community, the research agenda is set by the process of discovery. As new discoveries are made, theories are advanced, technology improves, and basic science continues to push forward with new discoveries. Problems are defined by finding shortcomings in the existing scientific knowledge base. The question at hand is whether the safeguards research agenda is set by technical innovation or political need. To answer this question, it is important to see how PNNL's safeguards research agenda is set from above and from within, and from political/policy and technical perspectives.

The International Atomic Energy Agency's (IAEA's) Department of Safeguards' purpose is to implement safeguards. In 2013, the Department developed short-, medium-, and long-term research and development plans to inform Member States of their current and future research and development requirements (IAEA 2013). Their research agenda is set by the technical and policy needs of the Department (IAEA 2013, IAEA 2014b). The IAEA has their own R&D capabilities, but still relies on monetary and R&D support through Member State Support Programmes. By signaling their priorities, the IAEA can influence Member States' safeguards policy and technical research agendas.

The U.S. government might take research cues from their contacts at the IAEA, but the President and Congress set the United States' research agenda when they pass a budget. The President and Congress take a host of factors into consideration when passing a budget, particularly the budget justification requests submitted by DOE and the U.S. Department of State (DOS). The DOE/NNSA's Defense Nuclear Nonproliferation FY 2015 Budget Request is driven by the imperative for U.S. leadership in nonproliferation initiatives both at home and abroad, but also by science and technology development that benefits the programs of the sponsoring agencies (DOE 2014). For FY 2015, President Obama allocated $27.9 billion in discretionary funds to the DOE and highlighted nonproliferation research as a priority (The White House 2014). The DOS' FY 2014 budget was driven by the need to support the IAEA's high priority safeguards projects (DOS 2014). For FY 2008, over $15 million was allocated to the State Department for the U.S. Support Program (USSP) to the IAEA (Nock and Hoffheins 2008). DOE and DOS are then in charge of allocating that money internally and to external recipients like PNNL. In the case of the USSP, the International Safeguards Project Office (ISPO) at Brookhaven National Laboratory (BNL), which manages the USSP, responds to IAEA requests and allocates tasks to relevant U.S. companies and national laboratories. This interaction between U.S. government entities is how policy and politics drive technical and policy research, but innovation and discoveries is another motivator.

The technical research agenda is set by innovation and the sharing of discoveries. This takes place during the numerous working groups, workshops, seminars, and conferences on technical topics held annually and hosted by the IAEA, Member States, and third parties both domestically and internationally. For example, in 2007, the IAEA and Japan Atomic Energy Agency co-hosted the Workshop on Advanced Safeguards Technology for the Future Nuclear Fuel Cycle (IAEA 2014a). The IAEA also may facilitate cooperation and share ideas through their Member State Support Programs. The INMM is a professional society, which hosts an annual meeting with sessions dedicated to new developments in safeguards technologies. These activities are not just limited to technical experts, but can also include policy researchers. Cooperation and communication is essential to innovation in both fields and the research ideas are articulated by the researchers themselves and then carried forward by colleagues.

PNNL's policy and technical researchers fund their projects through proposed work, which is where they and their work can influence the research agenda. In both cases, their findings might influence their proposals, but funding also has a role to play. DOE/NNSA and BNL's ISPO offer guidelines for proposals based on Department, Presidential, IAEA, and other research priorities and discoveries. PNNL employees individually and in groups, submit proposals to DOE/NNSA and/or DOS for approval. While PNNL employees can submit any proposal they like (within reason), they are constrained by the topics DOE/NNSA and DOS will approve. Essentially, all research at PNNL requires the authorization of funding, but it is up to the PNNL researchers themselves to come up with novel ideas. It is therefore a question about what drives researchers' interests. The next few paragraphs look at peer-reviewed articles that discuss science-policy relations to gain a better understanding of what the innovation drivers of policy and technical researchers are and why they have differing views.

The Oxford dictionary defines science as the systematic study the physical and natural world through observation and experiment (Oxford Dictionaries 2014). In this report, technical and scientific are used interchangeably to describe individuals whose research is based on physical science (i.e., physics, nuclear engineering, chemistry, computer science, etc.). Although there are nuances between applied and fundamental sciences and scientists, this paper does not make a distinction; further research accounting for the differences may be warranted. Policy, on the other hand, is the response by an authority to a problem (Birkland 2011). Therefore, a policy-oriented researcher investigates the effects of these responses on the affected population. Policy and technical researchers are not only divided by what they research, but also by their backgrounds.

Policy researchers and scientists are divided by cultural/educational factors, wider management and planning needs, and ideological preferences (Smith and Kelly 2003). The culture of scientists typically includes specialized training, which leads to long-term projects that are testable and reproducible, whereas policy usually places emphasis on a cross-disciplinary understanding and short-term goals with reactive decisions while taking into account the political environment including the changing priorities of the public and society as a whole (Smith and Kelly 2003). The division between policy researchers and scientists is perpetuated by perceptions of each other's research methods. The view of science as "a source of certified, neutral knowledge uncorrupted by the influence of politics" further divides the two communities by placing policy research in subsidiary category as both subjective and unmethodical (Smith and Kelly 2003). Furthermore, scientists generally believe that decision-making would be far more efficient using objective, rather than subjective, reasoning (Larason Schneider and Ingram 1997). Yet, scientific knowledge is just one of many considerations taken into account in policy research, others of which may be more subjective (i.e., public opinion, economics, law, etc.) (Zandbergen and Petersen 1995). The perceived subjectivity of one side and the objectivity of the other create an unbalanced paradigm where science is "fact" and policy is "conjecture." When in reality, some policy researchers use scientific methods and scientists often make subjective interpretations based on the opinions of the communities (political, religious, educational, etc.) of which they are a part (Cortner 2000). The disagreement among PNNL policy and technical researchers may be influenced by this difference in self-perception.

Despite the cultural differences between policy researchers and scientists, the future and forward momentum (essentially innovation) are what unites science and policy and drives them forward (McFadden et al. 2009). McFadden et al. also argues that both the management and understanding of change is a principal concern across both the science and policy fields (McFadden et al. 2009). Tomczak contends that science changes and develops with the needs of society and for that matter policy, although society and policy are not the only factors. With the focus on science-policy interface, the interaction is inherently cyclical and the relationship is reliant on one another, because science evolves to meet the needs of policy, and policy must evolve to the changes in science (Tomczak 2015). We argue that there is a "chicken-and-egg-like" relationship between scientific and policy development; it is unclear which came first, but the relationship is clear. In practice, individuals within those spheres (i.e., scientists and policy researchers) are largely self-isolating because of their backgrounds and perceptions, so it is unclear which has more influence on the other if any. This highlights a gap in the current literature, because there is no definitive conclusion as to which party—policy researchers or scientists—drive research.

3.2 Methodology

We looked at the publication dates of policy publications in comparison with technical publications to determine whether there is a trend over time of one category preceding or leading the other. We hope to see, for example, that a policy product was followed up by a technical product in the next time period, or vice versa. This method necessitated that we use the policy-technical coded SPARK dataset. To test for sensitivity, we used the subset that included products containing the word "safeguards." This information was then exported to R (R Core Team 2014) for analysis. R is programming language and software environment for statistical computing and graphics.

The product publication dates range from September 1, 2005 to August 31, 2013. The dates were chosen to ensure a representative sample from both before and after the inception of NGSI. The products each have a precise month, day, and year association. We coded them by month and assigned an adjusted fiscal year (September 1 to August 31).

The dataset was limited to three product types: abstracts, presentations, and reports because they have the most complete entries in ERICA[1]. Abstracts[2] and presentations[3] make up the largest proportion of the dataset and are most representative. Reports[4] represent a milestone in a project, i.e., a deliverable or a final report. They can be more indicative of a project's true timeline than presentations or abstracts; however, they are published less frequently. We conducted the analyses by product type, because no product type is all-inclusive.

3.2.1 Hypothesis

The literature does not indicate who would be most likely to drive research at PNNL, scientists or policy researchers. This analysis will test two alternative and corresponding null hypotheses:

H^a_1: Increases in policy product publications are correlated with an increase in technical product publications.

H^a_0: Policy research at PNNL is uncorrelated with the number of technical research publications.

H^b_1: Increases in technical product publications are correlated with an increase in policy product publications.

H^b_0: Technical research at PNNL is uncorrelated with the number of policy research publications.

Failing to reject both null hypotheses would suggest that no relationship exists between technical and policy research as a driver for the other, but further analysis using different methods would be needed to confirm the finding.

3.2.1.1 Methodology for the Analysis

Regression was used to test the hypotheses in order to control for multiple factors and predict technical/policy products by product type. To address biases in the data, several regression models were run before settling on the following specifications. The regression set one product (e.g., policy) as the dependent variable ($Products^a$) against the alternate (i.e., technology) as the independent variable ($Products^b$) and vice versa. To determine if policy products predict technical products and vice versa, the independent variable was lagged between k = 1 to 36 months in increments of 3 months, where "t" equals time. Fixed effects and a month variable (date the dependent variable [product] was published) were used to control for seasonality. To control for temporal dependence, the dependent variable was included as a

[1] Parker, Tomiann. 2014. Interview of Tomiann Parker (Pacific Northwest National Laboratory) by A Sayre (Pacific Northwest National Laboratory) "Information Release Discussion," 2 July 2014.

[2] An abstract is generally a brief summary of a larger, more comprehensive item, usually submitted to a conference or journal to determine interest in more information on the subject. When ERICA was first created, there was the assumption that there is a 1 to 1 ratio of abstracts to presentations. So for several years ERICA had the ability to link an abstract to a presentation- hence the existence of the abstract/presentation. It was determined that the 1 to 1 ratio was false, therefore the linking ability was removed and two independent categories were created (abstract and presentation).

[3] A presentation is a poster session product, PowerPoint or viewgraph slides or other materials used as the basis for lectures, conference talks, speeches, materials for technical courses, seminars, workshops etc.

[4] Reports are documents used to communicate the results of research, policy recommendations etc. often specified as contract deliverables and include the standard technical reports, Cooperative Research and Development Agreement (CRADA) reports, dissertations and theses, environmental impact statements, Laboratory-Directed Research and Development reports and technical assistance reports.

lagged independent variable in the regression. The variable TimeSinceNGSI describes the interval from the initiation of NGSI to publication in months. Products dated before NGSI have time zero. The variable "i" refers to the vector of products (e.g., i=0 is $PolicyProducts_t = \beta 1Month_t + \beta 2FiscalYear_t + \beta 3TimeSinceNGSI_t + \beta 4PolicyProducs_{(t-k)} + \beta 5TechnicalProducts^b_{(t-k)}$).

The generalized regression equation is as follows:

$$Products^a_i = \beta_1 Month_{it} + \beta_2 FiscalYear_{it} + \beta_3 TimeSinceNGSI_{it} + \beta_4 Products^a_{(t-k)i} + \beta_5 Products^b_{(t-k)i}$$

3.3 Results

Tables 3.1 and 3.2 describe the results of the regression. Red indicates a negative relationship, green a positive relationship, and blank indicates the results were not significant ($p > 0.1$). The regression results are inconclusive with respect to testing the hypotheses. Due to the instability of the regression coefficient signs and the lack of significant coefficients, we fail to reject the null hypotheses that the dependent and independent variables are uncorrelated.

Table 3.1. SPARK Dataset – Policy and Technology

	DV	IV	1	3	6	9	12	15	18	21	24	27	30	33	36
Abstracts	Policy	Technology													
	Technology	Policy													
Presentations	Policy	Technology													
	Technology	Policy													
Reports	Technology	Policy													
	Policy	Technology													

DV = dependent variable
IV = independent variable

Table 3.2. Safeguards Subset – Policy and Technology

	DV	IV	1	3	6	9	12	15	18	21	24	27	30	33	36
Abstracts	Policy	Technology													
	Technology	Policy													
Presentations	Policy	Technology													
	Technology	Policy													
Reports	Technology	Policy													
	Policy	Technology													

DV = dependent variable
IV = independent variable

3.4 Discussion

The negative relationships between variables were unanticipated, and dominate the significant findings. Most of the negative regression coefficients occur at 21-month (6 occurrences) and 1-month (4 occurrences) lag times. They might indicate missing variables in our regression. Alternatively, they could be causal if there is a substitution effect where publication in one category leads to increased funding in that category at the expense of the alternative. For example, if policy work led to innovation and drew additional resources away from technology research in the future, it would have a negative impact. This could indicate that the positive relationships rely on long-term influence rather than short-term reactionary stimuli. However, more research and additional variables would be needed to confirm this relationship.

The TimeSinceNGSI variable had a stabilizing effect on the outcomes of the regression. It removed two significant negative relations, but did not shift them towards positive significance. This would suggest that NGSI has some critical impact on the model by reducing potential omitted variable bias. However, it still underscores the need for a more complete regression equation. One problem could lie with the time variable we used for the time sequence. Publication date might be an incomplete indicator of what is driving research at PNNL. Essentially, the variable times are not normalized because they can describe the beginning, middle, or end of a project. Future research could consider other time-indicators, such as funding allocation dates, to normalize the time variable.

Furthermore, it is clear that NNSA determines the direction of research on a broad scale because they allocate funding to the whole NGSI program and all participating facilities. Thus, a research product (and findings) at one laboratory may drive funding of a complementary project at another. While there is some inter-laboratory collaboration, the NNSA has the broadest understanding of the scope and direction of safeguards research and development conducted at the DOE/NNSA contractor sites. This could explain why product submissions at PNNL have little relation, and this supports the expansion of the SPARK dataset. Future analyses with a comprehensive SPARK dataset may yield more significant results.

4.0 Conclusion

This report fulfills the Task 2 deliverable for this project by curating a document library of PNNL resources and training materials for input into the SPARK website. This report lays the basis for the intended SPARK dataset and its envisioned analysis capabilities. It started by describing the development of the dataset, including the methodology and scope. This process highlighted the need for an automated system. We also demonstrated how this dataset could be coded and subset to aid in research and analysis.

We also provided an example analysis to demonstrate the applicability of the envisioned capabilities. While the results of that analysis were inconclusive, they did highlight the need for an expanded SPARK dataset. This would help researchers understand and share information on the full scope of safeguards activities throughout the United States and abroad.

SPARK is not intended to draw any final conclusions about safeguards research, but could be a resource for researchers from both policy and technical fields to learn more about safeguards topics. A broader SPARK dataset may help draw new connections within and outside the safeguards focus. The keywords/theme development, although initially describing safeguards could be adapted to help researchers discover new information on other topics.

Furthermore, this project demonstrated that NGSI has and continues to make an impact on safeguards publications and research. In continuing the SPARK project, it is important to remember that NGSI was initiated to prepare the next generation of safeguards experts and capabilities. This expertise is not generated or sustainable without a robust institutional memory to contain and share the lessons learned and innovations. SPARK fills this gap in the safeguards research and will continue to do so as it is expanded and used. Pending funding, the next deliverable for Task 3 will develop the complete SPARK website for demonstration with this dataset.

5.0 References

Birkland TE. 2011. *Introduction to the Policy Process: Theories, Concepts, and Models of Public Policy Making*. Taylor and Francis, London and New York.

Cortner HJ. 2000. "Making science relevant to environmental policy." *Environmental Science & Policy*, 3(1):21-30. DOI:10.1016/S1462-9011(99)00042-8.

De Boer C and AS Velthuijsen. 2001. "Participation in Conversations About the News." *International Journal of Public Opinion Research*, 13(2):140-158.

DOE - U.S. Department of Energy. 2014. *Department of Energy FY 2015 Congressional Budget Request*. DOE/CF-0098, Volume 3. Available at http://energy.gov/sites/prod/files/2014/04/f14/Volume%203.pdf.

DOS – U.S. Department of State. 2014. *Congressional Budget Justification Volume 1: Department of State Operations Fiscal Year 2014*. Available at http://www.state.gov/documents/organization/207266.pdf.

Doyle JE. 2008. *Nuclear Safeguards, Security, and Nonproliferation: Achieving Security with Technology and Policy*. Butterworth-Heinemann, Oxford.

IAEA-International Atomic Energy Agency. 2002. *IAEA Safeguards Glossary, 2001 Edition.* International Nuclear Verification, Series 3, International Atomic Energy Agency, Vienna, Austria.

IAEA-International Atomic Energy Agency. 2010. *International Target Values 2010 for Measurement Uncertainties in Safeguarding Nuclear Materials*. STR-368, International Atomic Energy Agency, Vienna, Austria.

IAEA-International Atomic Energy Agency. 2013. *IAEA Department of Safeguards Long-Term R&D Plan, 2012-2023*. STR – 375, International Atomic Energy Agency, Vienna, Austria.

IAEA-International Atomic Energy Agency. 2014a. "JAEA-IAEA Workshop on Advanced Safeguards Technology for the Future Nuclear Fuel Cycle." 13-16 November 2007. Accessed November 15, 2014, at http://www-pub.iaea.org/mtcd/meetings/Announcements.asp?ConfID=1073.

IAEA-International Atomic Energy Agency. 2014b. "Safeguards." Accessed September 23, 2014, at http://www.iaea.org/safeguards/index.html. Last updated September 16, 2014.

Larason Schneider A and H Ingram. 1997. *Policy Design for Democracy*. University Press of Kansas, Lawrence, Kansas.

McFadden L, S Priest, C Green, and A Sandberg. 2009. *Basic Principles of Science and Policy Integration*. Spicosa Project Report, Flood Hazard Research Centre, Middlesex University, London.

Merriam-Webster. 2015. "agenda." Accessed January 6, 2015 at http://www.merriam-webster.com/dictionary/agenda.

NNSA - National Nuclear Security Administration. 2009. "NNSA Next Generation Safeguards Initiative." Accessed July 8, 2014 at http://nnsa.energy.gov/mediaroom/factsheets/nextgenerationsafeguards.

NNSA - National Nuclear Security Administration. 2010. "Next Generation Safeguards Initiaitve: Human Capital Development Program." Accessed July 8, 2014 at http://www.iaea.org/safeguards/symposium/2010/Documents/PPTRepository/108P.pdf.

NNSA - National Nuclear Security Administration. 2012. "Next Generation Safeguards Initiative Attracts University Students to Nonproliferation." Accessed July 8, 2014, at http://nnsa.energy.gov/blog/next-generation-safeguards-initiative-attracts-university-students-nonproliferation.

NNSA - National Nuclear Security Administration. 2014. "The Office of Nuclear Safeguards and Security." Accessed July 8, 2014, at http://nnsa.energy.gov/aboutus/ourprograms/dnn/nis/safeguards.

Nock C and B Hoffheins. 2008. *The U.S. Support Program to the IAEA - How it Works*. BNL-81294-2008-CP, Brookhaven National Laboratory, Upton New York.

Oxford Dictionaries. 2014. "science." Accessed September 23, 2014 at http://www.oxforddictionaries.com/definition/english/science.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. Available at http://cran.r-project.org/doc/manuals/fullrefman.pdf.

Schattschneider EE. 1975. *The Semi-Sovereign People: A Realist's View of Democracy in America*. Wadsworth, Boston and Cengage Learning.

Smith W and S Kelly. 2003. Science, technical expertise and the human environment. *Progress in Planning*, 60(4):321-394. DOI:10.1016/S0305-9006(02)00119-8.

The White House. 2014. "Department of Energy," *The Budget for Fiscal Year 2015.* Available at http://www.whitehouse.gov/sites/default/files/omb/budget/fy2015/assets/energy.pdf.

Tomczak M. 2015. "The precautionary principle." Blog entry dated July 9, 2007 accompanying lecture notes for Science, Civilization and Society. Accessed January 5, 2015 at http://sciencecivilizationandsociety.blogspot.com/2007/07/precautionary-principle.html.

Toomey C, J Olson, R Burbank, and M Madison. 2014. *SPARK Implementation Strategy*. PNNL- 23577, Pacific Northwest National Laboratory, Richland, Washington.

Wyse E and L Williams. 2012. "Next Generation Safeguards Initiative Attracts Young Professionals to Nonproliferation," NNSA Office of Nonproliferation and International Security Fall 2012 Highlights. Accessed February 5, 2015. http://nnsa.energy.gov/sites/default/files/nnsa/2013-10-29%20NIS%20Fall%202012%20Highlights.pdf.

Zandbergen P and F Petersen. 1995. “The role of scientific information in policy and decision-making.” *The Lower Fraser Basin in Transition: A Symposium and Workshop.* Kwantlen College, Surrey, BC, Canada.

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

U.S. DEPARTMENT OF
ENERGY

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

www.pnnl.gov