Thematic Shifts in Discussion and Questions About Consent-Based Siting[1]

Matthew D. Sweitzer,* Danielle N. Sanchez,* Alisa Rogers,* Thushara Gunda,* Lauren Drakopulos,* and Tara LaForce*

*Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185, msweitz@sandia.gov

INTRODUCTION

Nuclear energy production currently accounts for roughly 19% of the total energy generation portfolio of the United States (U.S.), makes up nearly 50% of the nation's carbon-free energy production, and results in approximately 2,000 metric tons of spent nuclear fuel (SNF) per year [1]. Thus, development and enactment of a strategy for managing commercial SNF remains a top U.S. priority to support ongoing nuclear- and climate-related objectives. In 2012, the Blue Ribbon Commission on America's Nuclear Future (BRC) issued a report to the Secretary of Energy that included recommendations for a new strategy to manage the back end of the nuclear fuel cycle [2]. One of these recommendations included the use of "a new, consent-based approach to siting future nuclear waste management facilities." Following the completion of the BRC report, the Department of Energy (DOE) issued its 2013 integrated waste management (IWM) strategy that adopted the BRC recommendation to use a consent-based siting approach to site SNF management facilities [3]. In 2015, DOE began seeking input from the public to design the consent-based siting process, and in 2017 DOE issued its draft consent-based siting process for federal consolidated storage and disposal facilities for SNF and high-level radioactive waste management [4].

The consent-based siting process is an approach to siting one or more federal interim storage facilities for SNF that centers the needs and concerns of people and communities. In implementing the consent-based sighting process, it may be helpful for DOE to seek out, understand, and consider ongoing or emerging public perceptions, conversations, questions, concerns, or interest in a timely fashion. To support DOE with this challenge, this report details a suite of natural language processing (NLP) data analysis tools that can assist with extracting relevant and timely information from a variety of texts. These tools have the potential to support more meaningful communication, outreach, and engagement efforts with a variety of members of the public, private industry leaders, local governments, Tribal Nations, and other stakeholders.

The purpose of this report is to demonstrate how NLP, in conjunction with qualitative analysis, can be used to better identify and understand key themes and questions present in texts provided by various individuals in response to solicitations for input on the consent-based siting process. Relatedly, the objectives of this report are two-fold: 1) illustrate how comment themes differ across six separate public solicitations, and 2) identify and categorize the technical subject matter of questions contained within the comments. These findings demonstrate the substantial utility of recent advancements in NLP techniques in summarizing and understanding a large body of public comments.

Public Solicitations

The collection and analysis of public input ensures that the consent-based siting process incorporates the necessary elements to foster an equitable and just siting process. The solicitations for public comment to date have covered a number of topics, including the design of the consent-based siting process and the role of private entities in the SNF management strategy. The solicitations within the comment response database (CRD) are organized into separate phases (PH).

The first two solicitations (PH1) concerned the draft consent-based siting process: CBS-PH1-PI was conducted between December 2015 and June 2016 and collected 653 unique comments from 634 commenters and in September of 2016 through the following month, CBS-PH1-DR gathered 56 comments from 69 unique authors. During PH2, solicitations were primarily focused on private initiatives and a second draft proposal. This phase included three solicitations: CBS-PH2-RFI which gathered 102 comments from 66 authors between October 2016 and January the following year, and CBS-PH2-DR and CBS-PH2-DR2[2] which collected 30 comments from 16 authors between January and April of 2017. Finally, PH3, consisting of a collection of public comments (CBS-PH3-RFI) from December 2021 to April 2022 and questions posed in an webinar (CBS-PH3-RFIWEB) conducted in October of 2022 gathered 16 and 128 comments from 11 and 38 unique authors, respectively. This most recent phase concerned the federal siting process and the then-upcoming funding opportunity announcement.

Responses are collected into the CRD via letter, email, fax, and phone call. In addition to hosting public correspondence, the CRD also provides a platform for organizing analysis outcomes (e.g., analysts' categorization schema applied to label comments) and archiving possible responses to the input received. The categorization of these comments by analysts into labelled groups is described further in the next section.

Prior Comment Analyses

After public input was received, a human analyst read the comments and labelled sections of them according to

[1]This is a technical report that does not take into account contractual limitations or obligations under the Standard Contract for Disposal of Spent Nuclear Fuel and/or High-Level Radioactive Waste (Standard Contract) (10 CFR Part 961).

To the extent discussions or recommendations in this report conflict with the provisions of the Standard Contract, the Standard Contract governs the obligations of the parties, and this report in no manner supersedes, overrides, or amends the Standard Contract.

This report reflects technical work which could support future decision making by DOE. No inferences should be drawn from this report regarding future actions by DOE, which are limited both by the terms of the Standard Contract and Congressional appropriations for the Department to fulfill its obligations under the Nuclear Waste Policy Act including licensing and construction of a spent nuclear fuel repository.

[2]This solicitation (82 FR 4333) had two phase assignments to indicate that separate sets of analyst labels were used in prior analyses to evaluate letters (former) versus emails (latter).

high-level subject matter categories present in the text using a guidance document which prescribed themes (called "binning guidance"). Human analyst-driven examination of the collected comments offers many benefits, including subject matter expertise that can provide a deeper understanding of critical contextual factors, as well as prepared responses.

Until the beginning of 2022, the binning guidance was pre-defined and differed between public solicitations (e.g., [5]). The most recent comment analyses used a binning guidance that was developed abductively (amending the pre-defined binning guidance through coding the first set of comments received; e.g., [6]), or was derived organically (i.e., inductively) as analysts reviewed the comments (e.g., [7]). Across all solicitations, consistent binning guidance was not used to analyze comments. Some pre-defined binning guidance documents were constrained in the subject matter the analysts were permitted to code for, whereas the latter two analyses were comparatively less constrained. This introduced substantive differences between the different solicitations, both regarding how subject matter was categorized and how the categorization process was implemented. This approach introduces complexity when attempting to draw inferences about the persistence or transience of different thoughts or concerns raised by commenters.

Study Objectives

As consent-based siting moves forward over the coming decade, gaining insights into the themes addressed in various types of public comments will be important, especially as comments reflect changes in the public's thoughts and concerns about SNF management over time. The challenges with consistently analyzing large collections of comments requires consideration of alternate methods [8]. For example, a NLP approach, which examines words and phrases within text to examine patterns and themes across a large body (i.e., corpus) of documents. Moreover, NLP can group relevant content more efficiently, reliably (i.e., perfectly consistent results), and adaptably (i.e., application to multiple corpora) than manual approaches, especially for larger bodies of text.

To evaluate the utility of NLP methods for supporting public solicitation analysis efforts, the research team developed and implemented a pilot approach using a single solicitation [9]. Specifically, a structural topic modeling (STM) approach was used to automatically group comments collected during the 2017 solicitation period (82 FR 4333) into subject matter categories (i.e., topics). The results indicated that, while there was a relatively high degree of correspondence between the variety of human analysts' and NLP-derived subject matter labels, the latter technique identified additional themes that were not present in the 2017 analysis (e.g., concerns about safety, the need for a mechanism to exit the consent-based siting process, and references to existing and proposed facilities; [9]). This report extends the prior NLP approach to a longitudinal analysis to examine all public input collected in response to solicitations between 2015 and 2022.

In addition to NLP, these objectives are supported by a qualitative approach to both generate labels for NLP topics as well as to capture information needs cited by commenters within the text. The latter is intended to support knowledge management efforts by identifying key concerns and relevant technical categories that those involved in community engagement activities for consent-based siting should be able to discuss with the wider public.

METHODS

Topic Modeling

This study used STM, which involves parsing a matrix of documents and terms, where the documents represent comments from the CRD, the terms reflect the words across the corpus of comments, and the cell values in the matrix reflect the word frequency of a particular term within a particular document. A topic modeling algorithm then identifies common terms to group comments with similar subject matter together. Specifically, compared to other topic modeling methods, the STM approach allows for the specification of document metadata as covariates [10]. For this analysis, several topic prevalence covariates were modeled, including the affiliation of author(s) and phase number. The STM was estimated using the *stm* R package [11]; see related newspaper content analysis [12] for additional details.

Results from the *searchk* procedure [11] indicated that $k = 24$ topics have optimal interpretability across all topics in the model. These topics (or groups of comments) that the STM identified were reviewed using additional NLP techniques (e.g., dendrograms and *estimateEffects*) as well as qualitative interpretation methods to generate additional insights. Dendrogram analysis uses topic-to-topic correlations within a document-by-topic theta matrix as inputs into a Euclidean distance function [12]. These topic-to-topic distances are plotted in a tree-like structure with hierarchical "branches" and "limbs" indicating the closeness of one topic to another; i.e., topics clustered together indicates a greater degree of co-occurrence across comments. This allowed for the identification of meta-topics (i.e., hierarchical groups of topics corresponding to broad subject matter overlap) that map to particular branches.

In addition to understanding relationships between topics, the relative prevalence of topics within the comments across public solicitations was also assessed. This was accomplished with *estimateEffects* plots that were generated by aggregating joint statistical estimates across topics contained in each meta-topic. This allows analysts to form estimates and confidence intervals (CIs) for the average meta-topic proportions for all documents with a specified category or value of the independent variable (i.e., document metadata; [11]).

Commenter Question Analysis

Abductive, manual content analysis informed the generation of questions that characterize public concerns and information needs. These questions aligned with the subject/object of concern cited in comments as well as the nature of that concern captured by subcodes and code specifiers. During this process, the concerns indicated in respondent comments were reframed as questions for which responses with answers could be generated. These questions were created to capture unique and nuanced concerns that were represented in each comment from the public. The abductive approach also facilitated the capture of unanticipated concerns and issues, allowing for a

more varied and complex understanding of public concerns and the identification of previously overlooked concerns.

Subject-matter categorization of questions followed a two-step procedure: first, one or more (i.e., aside from "process", categories were not mutually-exclusive) of five primary categories: *Basic Energy Generation*, *Storage*, *Transportation*, *Processes*, and *Impacts*. This top-level category was assigned based on the main subject of the question. Secondary question categories (i.e., subtopics) were then labeled when relevant. Because of the abundance of "process"-related questions identified in the first iteration of question categorization, questions were first categorized among the other four categories, and "process" was only used if no other category applied.

RESULTS

Topic Modeling

Among the 24 topics, subject matter labels were identified using top-fitting examples and a list of frequent and exclusive terms. International approaches to managing SNF (topic 16) was the most prevalent within the corpus, with respondents pointing to international examples as an avenue to inform the U.S.'s strategy. Community impacts (topic 12), which referred to the need for communities to have clear information on the benefits and risks of hosting a nuclear facility within their community, was the second most common topic. The third most prevalent topic refers to the role private companies and initiatives have in the generation and management of SNF (topic 6). Taken together, these topics are focused on the U.S.'s high-level strategy for managing SNF. Specifically, commenters point to learning from international examples and the rights, roles, and responsibilities of federal and private entities within this strategy. In addition, the "community impacts" topic highlights communities' desire for clear communication regarding both the benefits and risks associated with hosting a federal consolidated interim storage facility.

Next, the co-occurrence of topics within the STM were determined by analyzing the distance between topic scores. As indicated by the branch color scheme in Figure 1, five meta-topics (i.e., higher-order topics) were identified from the dendrogram. Moving clockwise around Figure 1, these topics include: *regionalized perspectives* (23.65%), *process considerations* (50.74%), *management & response* (12.99%), *transportation* (6.53%), and *equity* (6.09%).

Regionalized perspectives topics focus on procedural aspects of the consent-based siting process that have the potential to affect the commenter's local community, including technical and regional considerations for site selection, communicating benefits and risks to potential host communities, and examining international models of SNF management. Process considerations in comments largely refer to the need for greater transparency within the consent-based siting process. For example, comments in this meta-topic refer to the process and guidelines for identifying potential sites (topics 17 and 2), engaging with communities, and ensuring benefits and risks of hosting a facility are clearly communicated to potential host communities (topic 12). Management & response topics focus on the procedures necessary for managing SNF over time spanning the planning, management, and engagement

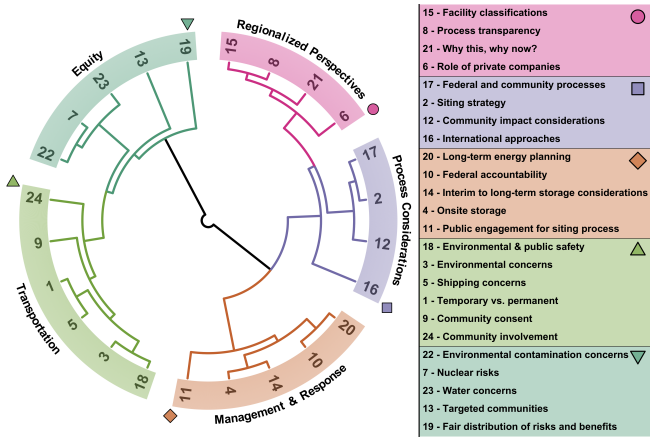

Figure 1. Dendrogram of topic-topic correlations across comments. Highlighted colors and shapes correspond to the five researcher-identified meta-topics.

necessary for building a procedural pipeline for storing SNF. Comments in the transportation meta-topic generally focus on the safety and security of moving SNF from the site of generation to a consolidated facility (interim or permanent) across different transport methods, including a focus on environmental concerns and other issues around transporting SNF. Finally, equity topics relate to the just and fair distribution of benefits and risks and the consideration of risks to the environment and access to clean drinking water. Notably, this meta-topic included discussion of the impact on underserved communities, including rural areas and Tribal Nations.

The authors next assessed the prevalence of the five meta-topics across comment solicitations through an estimate effects analysis (see Figure 2). Importantly, because the past comment analyses were not conducted using a common analytical tool, such comparative analyses were not possible. Across all solicitations, the meta-topic of process considerations was represented in the highest proportion of comments, becoming a dominant part of the discussion in latter comment solicitations compared to earlier ones (i.e., early 2017 onwards). The regionalized perspectives meta-topic was the next most prevalent, with the highest proportion of comments occurring in earlier comment solicitations (i.e., late 2016 and early 2017).

The remaining three meta-topics occur with much less frequency. Comments regarding equity and transportation meta-topics appeared with remarkably consistent frequency across all six public solicitations. The management & response meta-topic was the third most frequent meta-topic in late 2015, but its usage halved the following year and remained relatively low through the 2022 solicitations. Taken together, public comments about these subject matter appear to occupy a secondary tier of the public's interests to comments about process considerations and regionalized perspectives.

Commenter Question Analysis

236 questions were identified, with several duplicates spanning the five subject matter categories: *storage* (n = 76), including benefits and risks of being a host community, the technical requirements for being considered as a host site, and costs; *process* (n = 72), concerning the consent-based siting

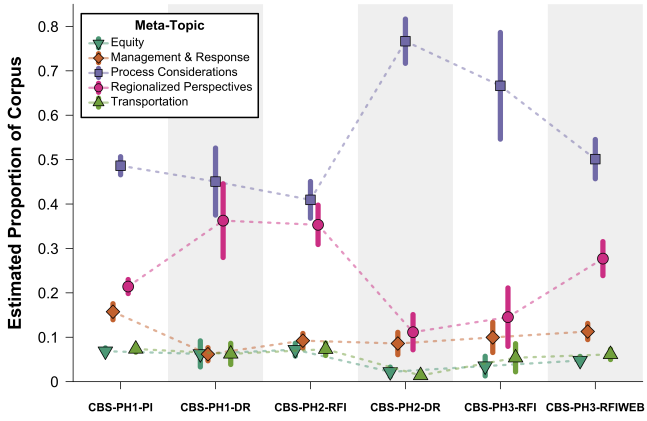

Figure 2. Estimated prevalence of meta-topics across solicitation periods. Solid vertical lines represent 95% confidence intervals of the estimates. The dashed lines have been provided to assist with trend detection – solicitation period is organized temporally along the x-axis, but is not a continuous variable (i.e., the temporal gaps between solicitations are not consistent).

process itself, and how risks would be mitigated for the community, public health, and the environment; *impacts* (n = 46) related to the environment, public health, and the long-term well-being of the potential host community; *transportation* (n = 29), including risks and safety concerns for communities located along transportation routes; and *basic energy generation* (n = 13) questions about technical information regarding nuclear science and about the benefits and risks of different renewable energy sources.

Additionally, 82 technical questions were identified from the master list of 236 questions. This list was further refined to reduce repetitive or similar questions, resulting in a final list of 37 technical questions. Frequently Asked Question responses aimed at public audiences were then developed by subject matter experts (SMEs) for a subset of 10 technical questions. An important lesson learned through this activity was both the importance, and the challenge, of translating technical concepts for public audiences. The team observed two sticking points: 1) differences in how the public and SMEs use key words and concepts, and 2) that technical concepts and questions related to SNF are often deeply embedded in, and dictated by, regulatory processes, which means SMEs face the challenge of translating both science and policy.

CONCLUSION

This report demonstrates the significant potential for NLP and other content-analytic techniques to gain insights into public narratives and concerns surrounding SNF management. These combined approaches could be used by DOE to enhance its analysis of public opinion, including how themes shift over time. Application of this approach to other sources of text, such as news media or communication materials, has the potential to offer additional insight regarding the implementation of the consent-based siting process.

ACKNOWLEDGMENTS

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan `https://www.energy.gov/downloads/doe-public-access-plan`. SAND2024-02398C

REFERENCES

1. DEPARTMENT OF ENERGY, "5 Fast Facts about Spent Nuclear Fuel," `https://www.energy.gov/ne/articles/5-fast-facts-about-spent-nuclear-fuel` (2022).
2. BLUE RIBBON COMMISION ON AMERICA'S NUCLEAR FUTURE, "Report to the Secretary of Energy," `https://www.energy.gov/ne/articles/blue-ribbon-commission-americas-nuclear-future-report-secretary-energy` (2012).
3. DEPARTMENT OF ENERGY, "Strategy for the management and disposal of used nuclear fuel and high-level radioactive waste," `https://www.energy.gov/sites/prod/files/Strategy%20for%20the%20Management%20and%20Disposal%20of%20Used%20Nuclear%20Fuel%20and%20High%20Level%20Radioactive%20Waste.pdf` (2013).
4. DEPARTMENT OF ENERGY, "Draft Consent-Based Siting Process for Consolidated Storage and Disposal Facilities for Spent Nuclear Fuel and High Level Radioactive Waste," `https://www.energy.gov/sites/default/files/2017/01/f34/Draft%20Consent-Based%20Siting%20Process%20and%20Siting%20Considerations.pdf` (2017).
5. R. KREUZER, T. MILEY, L. RODMAN, P. MIRICK, N. SARAEVA, and R. HOWARD, "Draft Consent-Based Siting Process Report – Summary of Public Comments." Tech. rep., PICS: NE Milestone No. M4IS-21PN0401092 (2021).
6. DEPARTMENT OF ENERGY, "Consent-Based Siting Request for Information Comment Summary and Analysis," `https://www.energy.gov/sites/default/files/2022-09/Consent-BasedSitingRFISummaryReport0915.pdf` (2022).
7. S. MOORE, "FOA Webinar: Informal analysis of question and comments shared." Tech. rep., Unpublished (2022).
8. M. SWEITZER and T. GUNDA, "Natural language processing in support of consent-based siting," Tech. rep., Sandia National Laboratories. PICS-NE Milestone No. M4IS-23SN010406062 (2023).
9. D. SANCHEZ, K. HACKETT, M. SWEITZER, and T. GUNDA, "Tag It: A Natural Language Processing Approach to Enhancing Delineation of Public Comments Regarding the Consent-Based Siting Process," Tech. rep., (SAND2022-11114 O). Sandia National Laboratories. PICS-NE Milestone No. M4IS-22SN010401154 (2022).
10. M. E. ROBERTS, B. M. STEWART, D. TINGLEY, C. LUCAS, J. LEDER-LUIS, S. K. GADARIAN, B. ALBERTSON, and D. G. RAND, "Structural topic models for open-ended survey responses," *American journal of political science*, **58**, *4*, 1064–1082 (2014).
11. M. E. ROBERTS, B. M. STEWART, and D. TINGLEY, "Stm: An R package for structural topic models," *Journal of Statistical Software*, **91**, 1–40 (2019).
12. M. D. SWEITZER and T. GUNDA, "Spatiotemporal Analyses of News Media Coverage on "Nuclear Waste": A Natural Language Processing Approach," *Nuclear Technology*, pp. 1–16 (2023).