

# Longitudinal Analysis of Public Input for Consent-Based Siting: A Case Study of Human-Machine Teaming

Thushara Gunda, Danielle N. Sanchez, Matthew D. Sweitzer, and Alisa N. Rogers

Sandia National Laboratories, Albuquerque, NM, USA

## Introduction

Nuclear energy plays a critical role in securing our nation's electricity supply, accounting for nearly 20% of annual United States (U.S.) production [1]. In addition to serving as a dependable baseload, nuclear energy is expected to serve as an important mechanism through which decarbonization goals for nation(s) are achieved [2]. Currently, the U.S. Department of Energy (DOE) is working to establish one or more federal consolidated interim storage facilities using consent-based siting as part of their overall Integrated Waste Management Program [3].

Consent-based siting seeks the willing and informed consent of people and communities to host a project in their area, and is designed to be flexible, adaptive, and responsive to community concerns. DOE has solicited input and feedback from the public since 2015 through different mechanisms (e.g., Invitation for Public Comment, Request for Information, or Request for Public Comments; Table 1). To augment our understanding of key themes from this content, which has been a time-consuming, manual, and solicitation-specific process to date, this study focuses on using human-machine teaming, specifically combining topic modeling with qualitative analysis, to generate longitudinal insights.

## Human-Machine Teaming

This study used topic modeling, which typically involves parsing a matrix of documents and identifying co-occurring common terms, to bin comments into groups with similar subject matter. These groups of comments were then reviewed using qualitative methods to generate additional insights for DOE.

Specifically, we used structural topic models which differ from other topic modeling methods, such as Latent Dirichlet Allocation, in that they allow for the specification of document metadata as topic covariates [4]. For this analysis, various covariates were included including affiliation of author(s) and Phase IDs (Table 1). The structural topic model was estimated using the *stm* R package [5].

Qualitative coding was then used to review the content of the comments for each topic and identify specific themes and concerns that were emphasized by the public. These insights were used to generate both topic labels and descriptions as well as identify recurring concerns and issues amongst public input.

Request No	Period of Solicitation		Title
(Phase ID)	Open Date	Close Date	
80 FR 79872 (CBS-PH1-DR)	12/23/2015	6/15/2016	Invitation for Public Comment to Inform the Design of a Consent-Based Siting Process for Nuclear Waste Storage and Disposal Facilities
81 FR 63475 (CBS-PH1-PI)	9/15/2016	10/30/2016	Request for Public Comment on the Draft Report Entitled 'Designing a consent-based siting Process: Summary of Public Input'
82 FR 4333 (CBS-PH2-RFI)	10/27/2016	1/27/2017	Request for Information on Approaches Involving Private Initiatives for Consolidated Interim Storage Facilities
81 FR 74779 (CBS-PH2-DR)	1/13/2017	4/14/2017	Request for Public Comment on Draft Consent-Based Siting Process for Consolidated Storage and Disposal Facilities for Spent Nuclear Fuel and High-Level Radioactive Wastes
86 FR 68244 (CBS-PH3-RFI)	12/01/2021	4/03/2022	Notice of Request for Information (RFI) on Using a Consent-Based Siting Process To Identify Federal Interim Storage Facilities.
No Request No (CBS-PH3-RFIWEB)	10/3/2022	10/3/2022	Consent-Based Siting Funding Opportunity Announcement Webinar (public questions)

Table 1. List of Public-Oriented Requests Analyzed

## References

- [1] U.S. Energy Information Administration (2022). "Nuclear explained." Accessed Dec 5, 2022 at <https://www.eia.gov/energyexplained/nuclear/nuclear-power-plants.php>
- [2] International Energy Agency (2019). "Nuclear Power in a Clean Energy System." Accessed Dec 5, 2022 at <https://www.iea.org/reports/nuclear-power-in-a-clean-energy-system>
- [3] U.S. Department of Energy (2021). "DOE Restarts Consent-Based Siting Program for Spent Nuclear Fuel, Requests Input on Interim Storage Process." Accessed Dec 5, 2022 at the URL specified: <https://www.energy.gov/articles/doe-restarts-consent-based-siting-program-spent-nuclear-fuel-requests-input-interim>
- [4] Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American journal of political science*, 58(4), 1064-1082.
- [5] Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). STM: An R package for structural topic models. *Journal of Statistical Software*, 91, 1-40.

## Topic Selection

For topic modeling, the number of topics ( $k$ ) used to group the content is specified by the analyst *a priori*. There are a variety of software tools to aid the analyst in determining an appropriate  $k$ -value. We used a procedure called "*searchk*", which provides averaged descriptive statistics of the models produced over various  $k$ -values [5]. These include semantic coherence (greater co-occurrence of high-probability words within the group of documents) and exclusivity (high-probability words don't occur in other topics), and relate to model interpretability.

An initial review of these statistics indicated that  $k=20$  and  $k=24$  were both promising topic numbers. A plot of per-topic statistics of these  $k$  values indicated a tighter grouping of the topics (i.e., smaller ellipsis) for  $k=24$  (see Figure 1). This signifies that the topics generated by  $k=24$  have more equivalent interpretability across all topics in the model and thus, selected for the final topic model estimation.

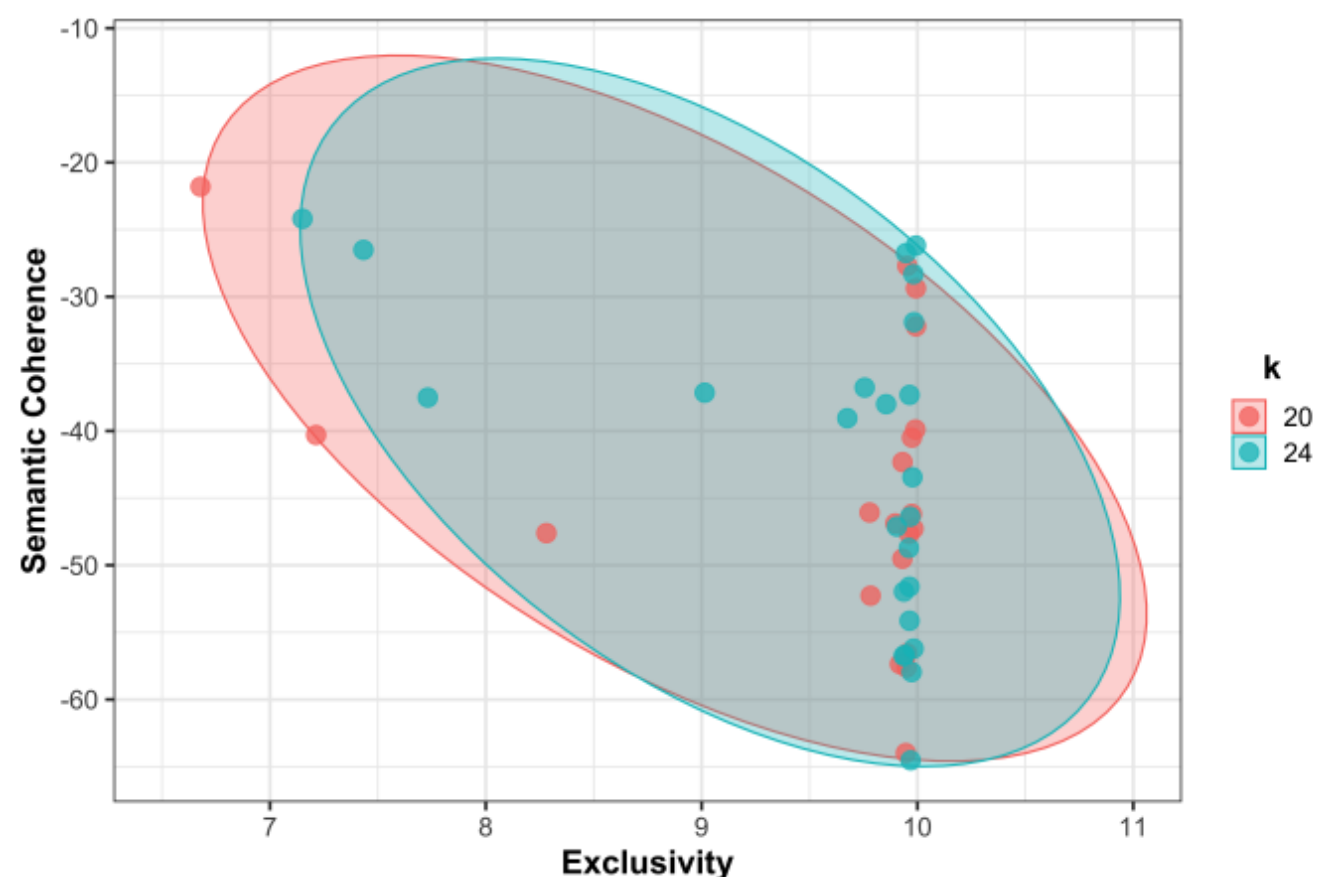


Fig. 1. Topic Descriptive Statistics by  $k$ -value

## Dendrogram Analysis

To interpret the meaning of each topic, the analyst can examine FRequently-occurring and EXclusive words (FREQX) in the topic, read high-scoring example comments, and use a dendrogram plot to investigate topics' co-occurrence among the documents. The latter procedure uses correlation values derived from the comment-to-topic fit values (i.e.,  $\theta$  matrix) to draw close relationships between topics that co-occur more frequently in a tree-like structure. The analyst can label "branches" of this graph by interpreting the topics that were grouped together to derive so-called "meta-topics".

As indicated by the branch color scheme in Figure 2, we identified five meta-topics from the dendrogram (clock-wise): **regional nuances** (23.65%), **process nuances** (50.74%), **management & response** (12.99%), **transportation** (6.53%), and **equity** (6.09%).

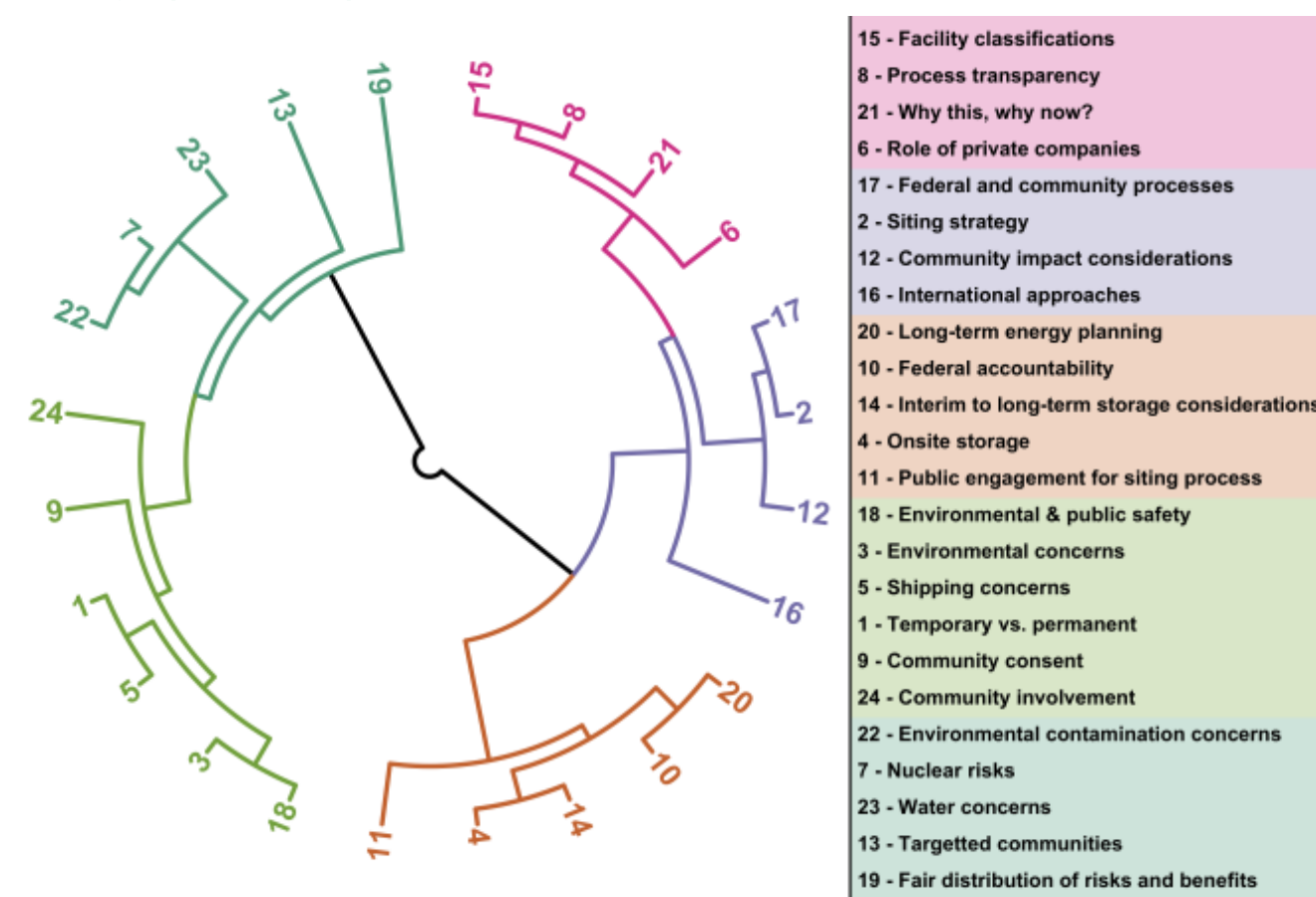


Fig. 2. Dendrogram of topic correlations across comments

## Topic Patterns

In addition to understanding relationships between topics, our analysis was also able to identify the relative prevalence of these topics within the public input. For example, a Sankey diagram demonstrates that siting process-related nuances dominated the primary topics, and there was significant overlap across the other (secondary) meta-topics (Figure 3). We also observed that, while the prevalence of most meta-topics remained fairly consistent across phases, process-related concerns increased in the three most recent calls for public input at the expense of regional concerns (Figure 4).

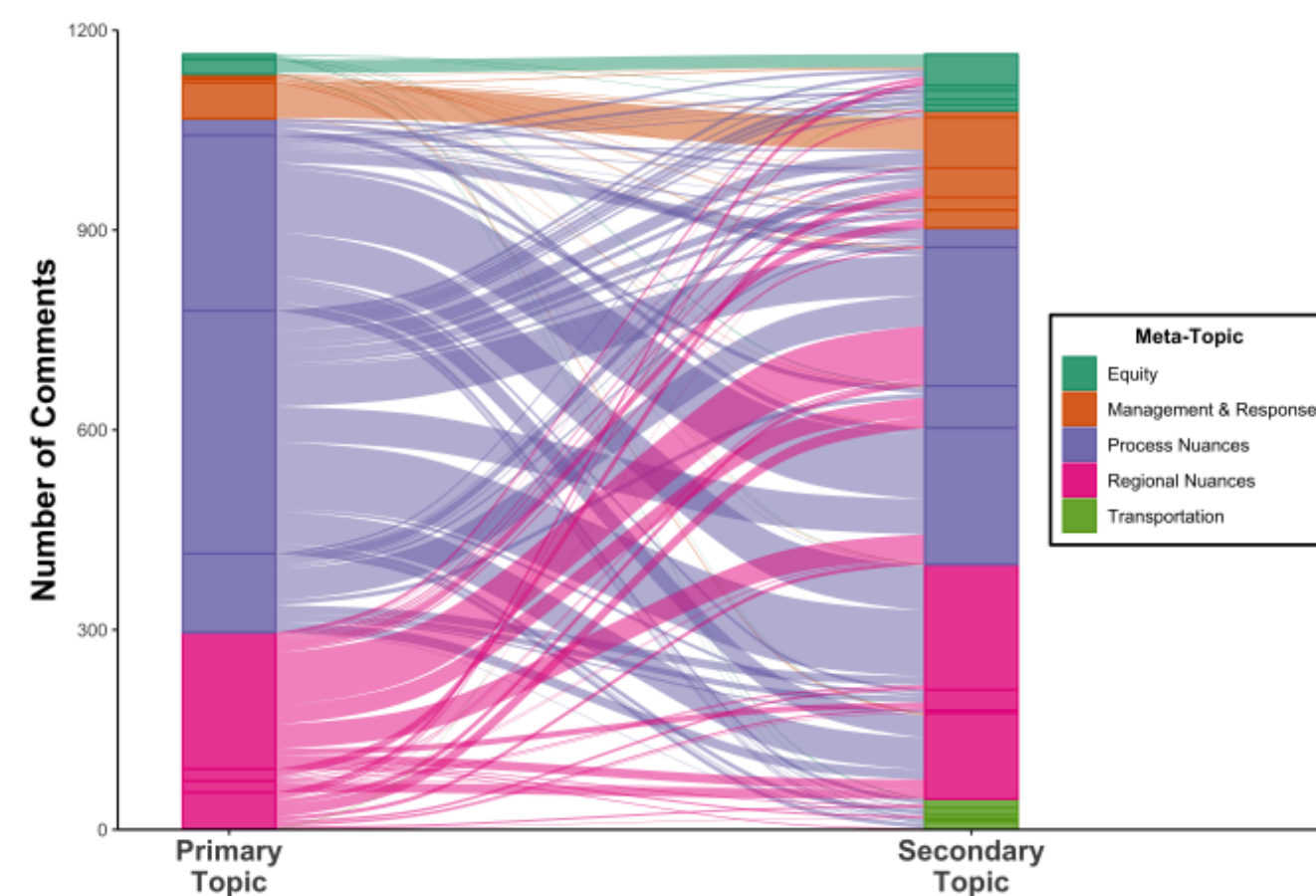


Fig. 3. Relationships between primary and secondary topic labels

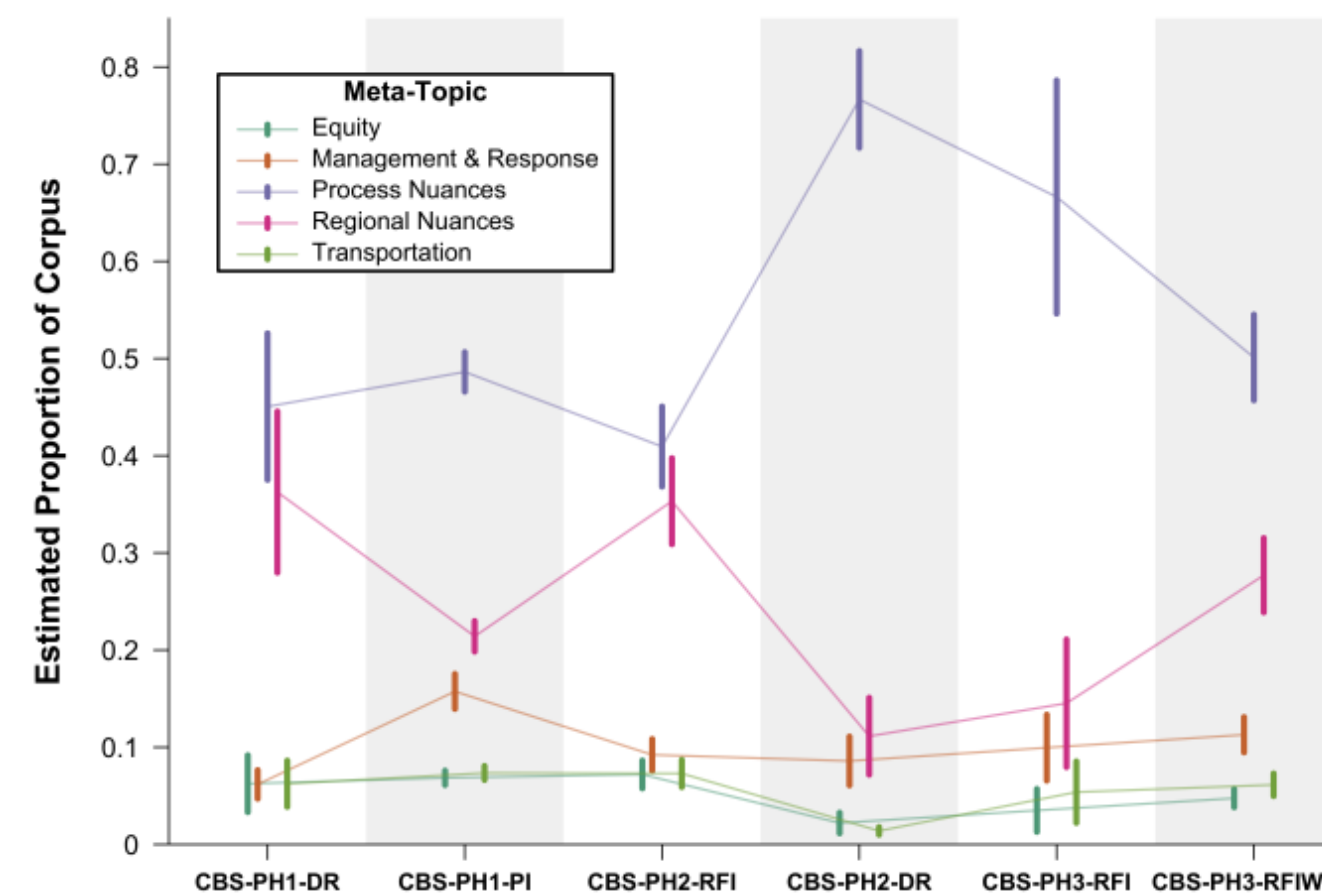


Fig. 4. Estimated prevalence of metatopics across phases with 95% CIs

## Qualitative Insights

In addition to generating labels, the qualitative analysis also identified a number of recurring questions that emerged in the public comments. Generally, these questions were centered around public understanding of:

- Value of nuclear energy (especially relative to other non-fossil energy sources);
- Risks associated with storing waste on-site vs. transportation (across natural and intentional threats); and
- Processes associated with engagement of stakeholders, including who provides consent.

In addition to varying levels of familiarity, these dimensions highlight the different nuances that DOE can integrate into consent-based siting communication approaches.

## Ongoing Work

The team is continuing to evaluate patterns in topics, such as priorities of certain topics amongst certain author affiliation types. These insights could help ensure the equitable development of resources for diverse groups.