Pacific Northwest
NATIONAL LABORATORY

Analysis of Peer Reviewer Comments for Energy Efficiency and Renewable Energy Research Projects using IN-SPIRE™

January 2023

David S. Gillen

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE)

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY

PNNL-33890

Analysis of Peer Reviewer Comments for Energy Efficiency and Renewable Energy Research Projects using IN-SPIRE™

January 2023

David S. Gillen

Prepared for
the U.S. Department of Energy

Pacific Northwest National Laboratory
Richland, Washington 99354

Acknowledgments

This report examines the strength, weakness, and recommendation themes from text analysis of 42,000 expert peer reviewer comments of R&D projects funded by the Office of Energy Efficiency and Renewable Energy (EERE) in the U.S. Department of Energy (DOE). It was prepared for DOE by the Pacific Northwest National Laboratory (PNNL). PNNL is operated by Battelle under Prime Contract No. DE-AC05-76RL01830.

Jeff Dowd of EERE's Strategic Analysis Office (SA) was the DOE Project Manager. Alex Maki, formerly AAAS Fellow in the EERE Strategic Analysis Office, assembled data for project.

The author extends appreciation to the following EERE staff who provided review comments of the draft study report:

- Jeff Dowd (EERE Strategic Analysis Office)
- Ammar Qusaibaty (BGS, LLC Contractor to the EERE Solar Energy Technologies Office)
- Brandon Rohnke (AAAS Fellow, DOE Office of Basic Energy Sciences)
- Randy Steer (EERE Information Technology Services Office)

Executive Summary

The Pacific Northwest National Laboratory (PNNL) used the IN-SPIRE™ visual analytics tool to identify recurring themes among comments noted during peer reviews of projects funded by the Office of Energy Efficiency and Renewable Energy (EERE) at the U.S. Department of Energy (DOE). The IN-SPIRE™ tool uses sophisticated text analysis and visualization to cluster similar content together and to identify reoccurring themes in large collections of text. As part of a pilot project IN-SPIRE™ was used to analyze 42,441 reviewer comments for 1,521 R&D projects across five EERE R&D offices – Advanced Manufacturing Office (AMO), Bioenergy Technologies Office (BETO), Hydrogen and Fuel Cell Technologies Office (HFTO), Vehicle Technologies Office (VTO), and Water Power Technologies Office (WPTO) from 2015-2019. The projects examined using IN-SPIRE™ included EERE-funded financial assistance, Small Business Innovation Research (SBIR), and different national laboratory projects.

This pilot IN-SPIRE™ analysis was performed across five review criteria commonly used in peer reviews performed by EERE offices. These are:

- Accomplishments and progress toward overall EERE goals,
- Approach to performing the work,
- Collaboration and coordination among partners and with other institutions,
- Relevance regarding alignment/support to EERE goals and objectives and to having meaningful/potential industry impacts, and
- Proposed future work such as having adequate plans, decision points, and future milestones

The analysis described in this report focused on identifying recurring themes within each criterion in three different concept areas: project strengths, project weaknesses, and recommendations by reviewers. For our analysis, we define themes as topics, ideas, and patterns of meaning that come up repeatedly. When possible, we interpreted each theme to inform an EERE R&D manager in how the theme can apply to EERE projects. For example, a strength theme might be that a project is well positioned for technology transfer, and a weakness theme might be that the accomplishments are not seen as relevant to industry partners. In total, 64 themes were identified. Some of these themes were unique to a particular criterion and concept, and some themes were repeated over multiple criteria and concept areas, although slight differences were observed.

Quantitative Results

The following is a high-level summary of the quantitative results.

- 45% of the reviewer comments contained at least one strength keyword, and 80% of the projects received at least one comment that included a strength keyword,
- The data shows a larger proportion of strength comments compared to weakness and recommendation comments.
  - More reviewer comments contained strength sentiment than weakness, by a factor of at least 2X.
  - For collaboration/coordination, the number of reviewer comments reflecting strengths was more than 4X the number of review comments containing a weakness.

  - For relevance, the number of reviewer comments reflecting strengths was more than 3X the number of review comments containing a weakness.

- 1,320 comments contained both a weakness keyword and a recommendation keyword. In many cases, reviewers were pointing out a weakness and also suggesting a recommendation to address this weakness, such as “Microreactors are notoriously difficult to transfer to market, and I would like to see the case for using them for this application.”

The analysis found 64 strength, weakness, and recommendation themes across the five review criteria, as indicated in Table ES.1.

Table ES.1. Counts of IN-SPIRE themes observed to occur in reviewer comments containing strengths, weaknesses, and recommendations for each criteria.

Review Criteria	Strengths	Weaknesses	Recommendations
Accomplishments	14 themes	13 themes	8 themes
Approach	14 themes	5 themes	9 themes
Collaboration/ coordination	13 themes	4 themes	6 themes
Relevance	8 themes	9 themes	7 themes
Proposed future work	8 themes	9 themes	7 themes

Section 3.0 of this report provides detailed quantitative results from analysis of these reviewer comments.

Lessons Learned from Previously Executed Projects to Help with R&D Management

The themes identified will be helpful to EERE R&D managers and can help inform their decisions about their future projects. Knowing the strengths and weaknesses of previous projects can provide the EERE R&D managers the information to guide future projects towards success and steer them away from failure.

Strength themes

Across the five review criteria, the following are the most occurring identified project strength themes. They are projects that have the following characteristics:

- Include cost assessment that provide information on cost efficiency and/or ways to improve cost efficiency of technology or an industrial process

- Create new materials and/or improve performance or development of materials

- Have potential for technology transfer and can scale/transition to industry to benefit industry and the market
- Exhibit good management
- Approach that has potential to overcome challenges and barriers
- Develop new models or improved on models from previous modeling activities
- Perform testing and especially real-world testing
- Engage collaborators – this includes having strong ties to industry, the right balanced mix of partners/collaborators, defined collaborator roles, and supplier partnerships and ties to consortium (where appropriate)
- Achieve or is on pace to meet goal and targets
- Adequately describe future work –includes useful applications for generated data, addresses scale up and/or increase in previous work, addresses previously identified challenges

R&D managers can look for these strengths in their projects or provide guidance to funded researchers to modify their project accordingly to better achieve these strengths.

Weakness themes

Across the five review criteria, following are the most occurring identified weakness themes. These are projects that have the following characteristics:

- Fail to address technical issues, barriers, and challenges
- Have limited technology transfer potential and face difficulty in transitioning to industry, or with marketability, or exhibit scaling challenges
- Exhibit missing or inadequate treatment of research aspects that experts consider important such as model development, gaps in application of materials used or developed, and data/measurement
- Are not well aligned to EERE goals or with accomplishments not relevant or seen as useful
- Have poor management such as with team composition or their interactions and coordination of team members
- Have a weak or flawed approach
- Fail to meet targets or goals or are likely unable to meet targets/goals

- Have inadequate industry connection, missing or poorly balanced mix of collaborators (academia, industry, DOE national labs)
- Demonstrate issues with cost assessment or cost modeling
- Have future plans that are too broad, lack focus, and/or do not address challenges in scaling from research to industry scales.

R&D managers should look for these issues in their projects and manage projects to avoid occurrence of these weaknesses.

Recommendation themes

Across the five review criteria, the most occurring identified recommendation themes include:

- Recommendations to strengthen cost assessment and/or include additional cost comparisons
- Improve data and modeling recommendations
- Further study of materials to include additional performance characteristics or use cases
- Expand testing to include additional materials and/or additional environmental conditions
- More performance data, performance analysis as compared to current methods and approaches
- Expansion of collaboratives – further engaging industry partners, expanded interaction with other partners, including engaging OEMs (where appropriate)
- Better articulate the potential impact of the R&D performed
- Expand or refine the approach or pursue focus on a particular area or direction
- Future work to include further validation or refinement of results for additional use cases

Opportunities to Further Improve the EERE Peer Review Process

The IN-SPIRE analysis has offered up suggestions for making future peer reviews more robust and enabling more valuable outcomes and interpretations of the results from the reviewers.

This IN-SPIRE analysis identified five opportunities for EERE to further improve its peer review process to enable reviewer comments to be more readily interpreted and used within the IN-SPIRE analysis framework or without the analysis.

- EERE could add an additional question in its guidance to elicit reviewer feedback directly related to the quality of the presentation materials, so quality-related kinds of weakness and recommendation comments can be responded to separately from core themes pertaining to the 5 primary review criteria.
- Offices could provide guidance to peer reviewers to encourage reviewers to include a judgement whenever they make an important observation. The guidance could also provide examples to illustrate the difference between observations without judgement and observations with judgement. For example:

  - This example makes an observation without making a judgement: “The planned tasks and sub-tasks have been completed or are on track for completion”
  - This example makes an observation and also includes a judgement: “The test data and progress is solid and illustrates that the activity is on track to continue with success”

- Explicitly ask the reviewers to identify strength, weaknesses, and recommendations for each criterion, and to provide distinct space for reviewers to provide this information. This will take the guesswork out of determining if a given comment should be interpreted as a strength or a weakness or a recommendation. An example format for soliciting this information is provided in Section 7.2

- Include more explicit guidance to the reviewers to identify the strengths and weaknesses more carefully in addition to positive and negative attributes of a project. For example, a comment like “This project has major problems. It should be drastically restructured or, preferably, canceled” indicates a negative viewpoint from this reviewer without including any specific weakness attributes that led to this viewpoint.

- Supply guidance to reviewers on how to encode multiple themes within a single comment, such as separating each theme in a bullet list, extra blank lines, or some other delimiter between individual themes. For example, rather than a reviewer writing “This project has engaged collaborators in industry and is working well with OEM manufacturers to reduce their production costs for materials that will exceed program targets”, the reviewer should split this sentence into three sections: one highlighting a strength in collaboration, one highlighting the impact on OEM manufacturers, and one highlighting exceeding program goals. These sections could be individual sentences, or they could be separate paragraphs within the same comment block. Specific guidance should specify how to separate these sections, such as using blank space between paragraphs.

Thematic analysis can inform more strategic decision making by EERE R&D managers through promotion of projects that exhibit strength themes, steering a portfolio away from weaknesses, providing project oversight/management to preempt occurrence of potential weaknesses, and being proactive about guiding projects based on recommendations given to previous projects. Through the analysis of over 42,000 reviewer comments across multiple EERE R&D programs, this IN-SPIRE analysis enables EERE R&D managers to learn lessons to aid project management based on information that extends beyond their own experiences.

Contents

Figures

Tables

1.0 Introduction

The Pacific Northwest National Laboratory (PNNL) used the IN-SPIRE™ Text Analysis tool to identify key themes from review comments made in response to evaluating R&D projects funded by the Office Energy Efficiency and Renewable Energy (EERE) at the U.S. Department of Energy (DOE). These themes identified strengths, weaknesses, and recommendations around five review criteria commonly used in peer reviews performed by EERE offices.[1] The five criteria are commonly used by the EERE offices include:

- Accomplishments and progress toward overall and DOE & EERE goals,
- Approach to performing the work,
- Collaboration and coordination among partners and with other institutions,
- Relevance regarding alignment/support to DOE goals and objectives and to having meaningful/potential industry impacts,
- Proposed Future work such as having adequate plans, decision points, and future milestones.

The identified strengths, weaknesses, and recommendation themes will be useful to EERE for distilling current insights that may also inform decisions around future project management and selection.

After analyzing 42,441 reviewer comments, PNNL identified 64 different themes contained within these reviewer comments. Of these, 23 were themes associated with strengths, 21 were themes associated with weaknesses, and 20 were themes associated with recommendations.

1.1 Data

PNNL was provided a spreadsheet from the EERE sponsor containing 70,670 reviewer comments of 2,208 projects. Each comment was in response to a project being evaluated in a given criterion during a peer review. The analysis evaluated projects from across five EERE R&D offices: Advanced Manufacturing Office (AMO), Bioenergy Technologies Office (BETO), Hydrogen and Fuel Cell Technologies Office (HFTO), Vehicle Technologies Office (VTO), and Water Power Technologies Office (WPTO) from 2015-2019. Each row in the spreadsheet represented an individual comment for a specific project and criterion, and contained various metadata fields for each comment, such as the EERE office, year of the review, project title, criterion, and actual comment text. The types of projects in the database were R&D, technology deployment, and analysis projects. The spreadsheet of comments was assembled by EERE, including the association of comments with the designated criterion and the labeling of projects as R&D, technology deployment or analysis projects.

[1] Peer Review Guidance December 2020 (EERE G 413.001). https://www.energy.gov/eere/analysis/peer-review-guidance-december-2020-eere-g-413001

PNNL performed its analysis on a subset of the reviewer comment spreadsheet file, focusing on R&D project reviewer comments in the 5 criteria listed above. Analysis project[1] reviews were removed from the spreadsheet, and reviewer comments for criteria other than the 5 criteria listed above were also removed. This reduced the reviewer comment spreadsheet to 42,441 rows, representing 1,521 R&D projects. The projects examined using IN-SPIRE™ included EERE-funded financial assistance, Small Business Innovation Research (SBIR), and different national laboratory projects.

1.2 IN-SPIRE™

IN-SPIRE™, powerful information visualization software developed by PNNL, can provide people the ability to identify new insights in the data they already have.

IN-SPIRE™ can quickly and automatically convey the gist of large sets of unformatted text documents such as technical reports, web data, newswire feeds, and message traffic. IN-SPIRE™ can handle real-time data by adding new documents as they get added. It also processes foreign language data and provides robust support for translation. By clustering similar documents together, this Windows-based software unveils common themes and reveals hidden relationships within the collection. IN-SPIRE™ allows analysts to spend more time exploring the information they find most relevant and less time sifting through the mass of irrelevant documents.

IN-SPIRE™ analyzes a corpus of text and determines key topics or themes in each to create a signature for each document in the collection. IN-SPIRE's two main visualizations display representations of the documents which have similar or related topics to appear closer. The Galaxy visualization uses the metaphor of the stars in the night sky with each star representing an individual document. The ThemeView™ visualization uses a three-dimensional terrain map display to provide a high-level overview of the data.

There are a few key analytics that contribute to IN-SPIRE's capabilities. System for Information Discovery[2] (SID) is an algorithm similar to Latent Semantic Indexing[3] (LSI) to perform topic modeling and creates the mathematical vector model for text content. Rapid Automatic Keyword Extraction[4] (RAKE) is used within IN-SPIRE to extract candidate keywords from text, including

[1] An effort was made to remove deployment and analysis projects from the spreadsheet. For analysis, the following are example areas where projects were removed from the spreadsheet.

- Techno-economic modeling
- Market trends analysis
- Environmental impact analysis
- Assessments to understand policy, regulatory, and market barriers
- Life-cycle analysis
- Risk assessment
- Potential for job creation

[2] System for Information Discovery (SID): http://www.google.com/patents/US6484168

[3] Latent Semantic Indexing (LSI): Deerwester, Scott, et al. "Indexing by latent semantic analysis." *Journal of the American society for information science* 41.6 (1990): 391-407.

[4] Rapid Automatic Keyword Extraction (RAKE): Rose SJ, DW Engel, NO Cramer, and WE Cowley. 2010. "Automatic Keyword Extraction from Individual Documents." Chapter 1 in *Text Mining: Application and Theory*, vol. 1, ed. MW Berry, J Kogan, pp. 3- 20. John Wiley & Sons, Chichester, United Kingdom. http://media.wiley.com/product_data/excerpt/22/04707498/0470749822.pdf

the identification and extraction of multi-word phrases. Computation and Analysis of Significant Themes[1] (CAST) is used to organize keywords into themes.

[1] Computation and Analysis of Significant Themes (CAST): Rose SJ, RS Butner, WE Cowley, ML Gregory, and J Walker. 2009. "Describing Story Evolution from Dynamic Information Streams." In IEEE *Symposium on Visual Analytics Science and Technology (IEEE VAST) VAST 2009*, Oct. 12-13, 2009, Atlantic City, NJ, pp. 99-106. IEEE , Piscataway, NJ. doi:10.1109/VAST.2009.5333437

2.0 Method

A summary of our method follows in the subsequent sections, highlighting our preprocessing and initial preparation of the data, and the tools and techniques we used within IN-SPIRE™ to achieve our results. Full details of our method, including definitions of terminology we use, can be found in Appendix A.

For our analysis, we define themes as topics, ideas and patterns of meaning that come up repeatedly. When possible, we interpreted each theme to inform an EERE R&D manager in how the theme can apply to EERE projects. Themes that appeared to be unrelated collections of keywords were discarded for this analysis.

2.1 Data Preparation

Before PNNL could begin the analysis process, we did some pre-processing of the data to make it more suitable to extract the insights we were looking for. First, we subsetted the data to exclude any non-R&D projects and remove reviewer comments not aligned to the 5 criteria. Second, we split comments so each sentence became a distinct record in IN-SPIRE™ for the analysis.

We subsetted the original dataset to only include comments from R&D projects. This reduced the number of records for analysis and reduced potential noise that was introduced by including non-R&D project comments. We also removed reviewer comments that were not associated with the 5 criteria we were focused on. After performing these removals from the spreadsheet, the number of projects being reviewed was reduced from 2,208 to 1,521, and the number of reviewer comments being reviewed was reduced from 70,670 to 42,441.

We also split individual comments into separate sentences, resulting in a separate IN-SPIRE™ record for each sentence. We did this because many comments contained multiple individual thoughts within a single comment, and we assumed these ideas would organize into more coherent themes if treated separately. The 42,441 reviewer comments were split into 110,309 records for analysis.

2.2 Approach

PNNL's approach to identifying and extracting recurring themes was to use the IN-SPIRE™ text analysis software. PNNL leveraged the sponsor's expertise and knowledge with the reviewer comment data to perform the preparation and curation of the dataset, allowing PNNL to focus on the analysis portion of the project.

PNNL approached this as an information discovery project. PNNL did not know ahead of time what kinds of themes existed in the data, and there was no taxonomy of themes available to try to map the dataset content into. PNNL applied a data-driven approach that leveraged IN-SPIRE's capabilities for information discovery.

PNNL also utilized an approach that avoided the need to read each and every reviewer comment. Since the goal of the analysis was to identify recurring themes, we utilized an approach that leveraged IN-SPIRE's ability to group and organize similar content together, as a guide to identifying the highly occurring themes within the datasets. When seeking themes of strengths, weaknesses and recommendations, we applied a general approach of "casting a

wide net" when seeking out these concepts of interest, and utilizing IN-SPIRE's capabilities to work through the results. While this approach may introduce false-positives, it also reduces the false-negatives.

2.3 IN-SPIRE™ Methodology

IN-SPIRE™ has many analytics and visualizations, but PNNL used some key functionality to identify key themes within reviewer comments. These included Lexical Analysis and the Galaxy visualization for the qualitative analysis, and Lexical Analysis, the Summary Tool, and the Facets tool for the quantitative analysis.

We used Lexical Analysis to identify records that are associated with strengths, weaknesses, and recommendations. This was done by creating custom lexicons of keywords associated with strength concepts, weakness concepts, and recommendation concepts, respectively. These lexicons evolved throughout our analysis process using an iterative process, including criterion-specific lexicons to capture subtle differences in conceptualization within a specific criterion.

We used the Galaxy visualization to explore clusters of records that IN-SPIRE™ organized together based on statistical and mathematic similarity in content. These clusters require manual exploration to identify if a theme can be identified to explain the underlying basis of IN-SPIRE™ clustering them together. We also considered the vocabulary and top terms of each cluster, and made iterative refinements to the vocabulary and stop words for each dataset.

Our methodology utilized a lot of iteration and trial and error to refine our vocabulary for both lexicon terms and Galaxy cluster terms. Lexicon term selection started with some keywords that we seeded from the specific concept plus some synonyms of the concept. To expand the lexicons, we utilized the Charged Terms view within IN-SPIRE to discover additional terms highly related to our lexicon terms. We also inspected the set of reviewer comments that didn't match our dataset and looked at term distributions to discover candidate terms to add to our lexicons.

For Galaxy terms, we would use the "outlier terms" function to remove individual terms that were influencing the clustering algorithm in ways that were undesirable. For example, we did not want clusters of comments around organization names, technology names, or other adjectives and verbs unrelated to our theme concepts. We would iteratively outlier terms, then allow IN-SPIRE to re-cluster the records to see how they would organize into new clusters. Our outlier terms would be applied as "stop major" terms for future iterations of these datasets.

A more detailed explanation of the method we used within IN-SPIRE™ can be found in Appendix A of this report.

3.0 Reviewer Records by Theme Area and Criteria

A total of 110,309 records were included in our analysis. These records were derived by splitting reviewer comments at sentence boundaries, so each sentence contained within a reviewer comment became a record for our analysis. Using this process, our records represented individual sentences extracted from 42,441 individual reviewer comments for 1,521 separate projects. These records were then split into three datasets: a subset of records containing strength keywords, a subset of records containing weakness keywords, and a subset of records containing recommendation keywords. In the descriptions below, "reviewer comment" refers to a complete comment written by a reviewer, and a "record" is an individual sentence extracted from a reviewer comment.

We processed each of these datasets through IN-SPIRE and recorded some high-level characteristics before beginning our identification of specific themes. IN-SPIRE provides raw counts of term frequencies, extracted themes using two built-in analytics: Rapid Automated Keyword Extraction (RAKE) and Computation and Analysis of Significate Themes (CAST), and counts of term frequencies for terms within our lexicons.

From the total 110,309 records, 24,972 of these records contained a strength keyword. These records originated from 19,050 original reviewer comments for 1,507 unique projects. We looked at term usage in these records and found the following characteristics:

- The top overall keywords in this dataset are approach, significant, progress, cost, and future
- The top themes identified by RAKE and CAST are "progress, accomplishments, goals, technical accomplishments", "approach, effective" and "future, plan"
- The most commonly used strength lexicon keywords were: good, significan* (including all suffixes, such as significant and significance), success* (including successful and successfully), accomplish* (including accomplishment), and excellent

From the total 110,309 records, 9,788 of these records contained a weakness keyword. These records originated from 8,042 original reviewer comments for 1,390 unique projects. We looked at term usage in these records and found the following characteristics:

- The top overall keywords in this dataset are challenges, approach, difficult, future, unclear
- The top themes identified by RAKE and CAST are "unclear, approach", "challenges, address, identified, barriers, remaining challenges", and "difficult, progress, evaluate, assess"
- The most commonly used weakness lexicon keywords were: "not clear", difficult* (including difficulty), concern* (including concerning and concerns), unclear, and lack* (including lacking and lacks)

From the total 110,309 records, 4,964 of these records contained a recommendation keyword. These records originated from 4,453 original reviewer comments for 1,248 separate projects. We looked at term usage in these records and found the following characteristics:

- The top overall keywords in this dataset are: expected, future, cost, suggest and approach
- The top themes identified by RAKE and CAST are "expected, approach, cost", "suggested, future, focus", and "suggest, reviewer"

- The most commonly used recommendation lexicon keywords were: “it would be”, suggest* (including suggests and suggestion), expect* (including expected and expecting), “I would” and recommend* (including recommending)

Table 1 summarizes the number of reviewer comments for each criterion, and the percentage of reviewer comments that contained a strength keyword, a weakness keyword, or a recommendation keyword.

Table 1. Percentages of reviewer comments in each criteria containing a strength, weakness or recommendation.

	# Reviewer Comments	% Strength	% Weakness	% Recommendations
Approach	9,287	50%	26%	11%
Accomplishments	8,357	63%	24%	13%
Collaboration	7,224	50%	12%	10%
Relevance	8,930	31%	11%	5%
Future Research	8.643	32%	20%	14%
Total	42,441	45%	19%	10%

Some observations from the data presented above include:

- 45% of the reviewer comments contained at least one strength keyword, and 80% of the projects received at least one comment that included a strength keyword,
- 19% of the reviewer comments contained at least one weakness keyword, and 73% of the projects received at least one comment that included a weakness keyword
- 10% of the reviewer comments contained at least one recommendation keyword, and 66% of the projects received at least one comment that included a recommendation keyword.
- 1,320 comments contained both a weakness keyword and a recommendation keyword. In some of these comments, we found reviewers were pointing out a weakness and also suggesting a recommendation to address this weakness, such as “Microreactors are notoriously difficult to transfer to market, and I would like to see the case for using them for this application.” In these cases, this comment was considered for both weakness themes and recommendation themes.
- The data shows a larger proportion of strength comments compared to weakness and recommendation comments.
  - More reviewer comments contained strength sentiment than weakness, by a factor of at least 2X.
  - For collaboration/coordination, the number of reviewer comments reflecting strengths was more than 4X the number of review comments containing a weakness.

  - For relevance, the number of reviewer comments reflecting strengths was more than 3X the number of review comments containing a weakness.
- The number of reviewer comments containing weakness themes was approximately 2X the number of comments containing a recommendation theme.

Among the reviewer comments, strengths were judged to be more dominate than weaknesses, particularly so for collaboration – although there were weaknesses found in all criteria areas. Several explanations could explain these results:

- Reviewers may have included more strengths than weaknesses or recommendations in the review comments.
- Use of word such as “good”.
- It can be challenging to detect negative concepts, like weakness, because in English, we often express negative concepts by negating a positive word.
- Our lexicons were very broad, perhaps overly broad.

We also observed that many comments contained mentions of multiple strength keywords, weakness keywords, and recommendation keywords, all within the same comment. The Venn diagram in Figure 1 shows the overlap among the usage of these keywords in both original reviewer comments and our individual records of analysis.

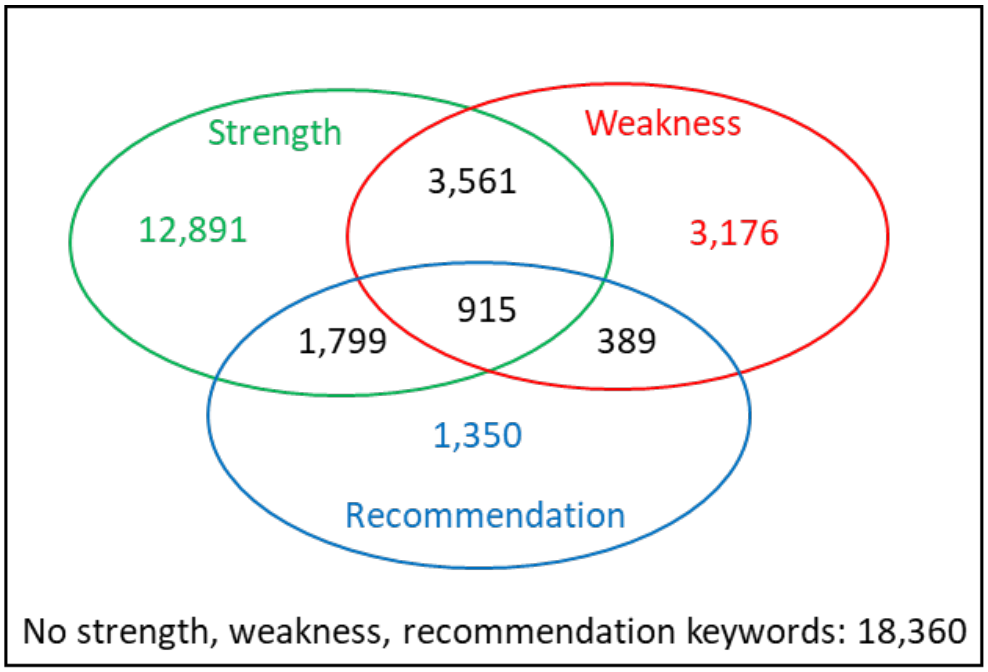

Figure 1. Venn diagram showing the counts of reviewer comments containing strengths, weaknesses and recommendations, and showing the overlap among these groups.

In all of overlapping cases, we considered the same comment for multiple concepts, so, for instance, the 2,714 comments that overlap the Strength and Recommendation bubbles (1,799 + 915) were considered for both strength themes and recommendation themes. The 915 comments in the middle were considered for strengths, weaknesses and recommendations.

Some examples of records that included all three types of keywords:

- "*Success* would be *valuable* even if it could not be shared in a lot of other applications – It is *not clear* why we should believe that there could be multiple applications of the *expected* outputs."
  - Strength keywords: success, valuable
  - Weakness keywords: not clear
  - Recommendation keywords: expected
- "The decrease of industrial usage is *concerning*, and *it would be valuable* to understand why and if it is a *significant*, sustained trend"
  - Strength keywords: valuable, significant
  - Weakness keywords: concerning
  - Recommendation keyword: it would be
- "The *expected* cost of the carbon capture system was not discussed explicitly, and one possible *concern* is that once designed, the maintenance of the system might be a *significant* expense"
  - Strength keywords: significant
  - Weakness keywords: concern
  - Recommendation keyword: expected

We also observed a substantial number of reviewer comments did not contain any keywords from our strengths lexicon, weaknesses lexicon or recommendations lexicon. We studied this collection of reviewer comments to make sure we were not excluding any major themes. We found this collection largely contained statements where a reviewer was stating a fact, or a statement where a reviewer was offering an observation without making judgement. If the reviewer intended to highlight a strength or weakness, or make a recommendation, the language that was used was either too subtle, or too infrequent for us to discover and add to our lexicons. These reviewer comments were comprised of approximately 43% of the overall comments, and these comments were omitted from the remainder of our analysis process.

4.0 Strength Themes

IN-SPIRE™ was used to identify themes associated with project strengths across the five criteria of Accomplishments, Approach, Collaboration and Coordination, Relevance, and Future Research. The strengths within each criterion were identified separately and presented below in Sections 4.1 through 4.5.

Across the different criteria, there were many recurring strength themes:

- Projects that were cost efficient and/or improving the cost efficiency of an industrial process
- Projects that were creating new materials and/or improving on material development
- Projects with strong ties to industry benefits or having the potential for benefiting an industry
- Projects with a strong management approach, with well-defined milestones and measurable goals and objectives
- Projects that clearly identified barriers and had approaches for overcoming or addressing them
- Projects that developed or improved on models and modeling activities
- Projects that performed testing, especially real-world testing
- Projects that had an engaged collaborators with well-defined roles for each collaborator
- Projects that aligned well towards EERE goals and performance targets.

The following subsections 4.1 through 4.5 present detailed result findings for strength themes by each of the five review criteria. Each subsection presents interpretable strength themes for each review criteria and theme percentages relative to the total number of records containing a strength keyword. Note that in many cases, a reviewer may have used a strength keyword in a comment, but the keyword was used in the absence of an interpretable strength theme. This resulted in theme percentages that only cover a portion of overall records and reviewer comments. More explanation of this is provided in the Discussion Section 7.0 of this report.

For each identified theme, we present excerpts taken from specific reviewer comment records. We do not define any of the referenced acronyms in the example reviewer comments because they are direct quotations from peer review records. The names of reviewers, EERE offices, consortia, and institutions have been omitted from the excerpts and replaced with [ ] in the reported records.

4.1 Detailed Results of Accomplishment Strength Themes

This section contains the detailed results of IN-SPIRE's analysis of strength themes contained in Accomplishment comments. This dataset contained 6,633 individual records, extracted from 4,637 reviewer comments for 1,258 different projects. Statistically, IN-SPIRE™ organized these records into 81 clusters, which distilled to 14 interpretable themes. Figure 2 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

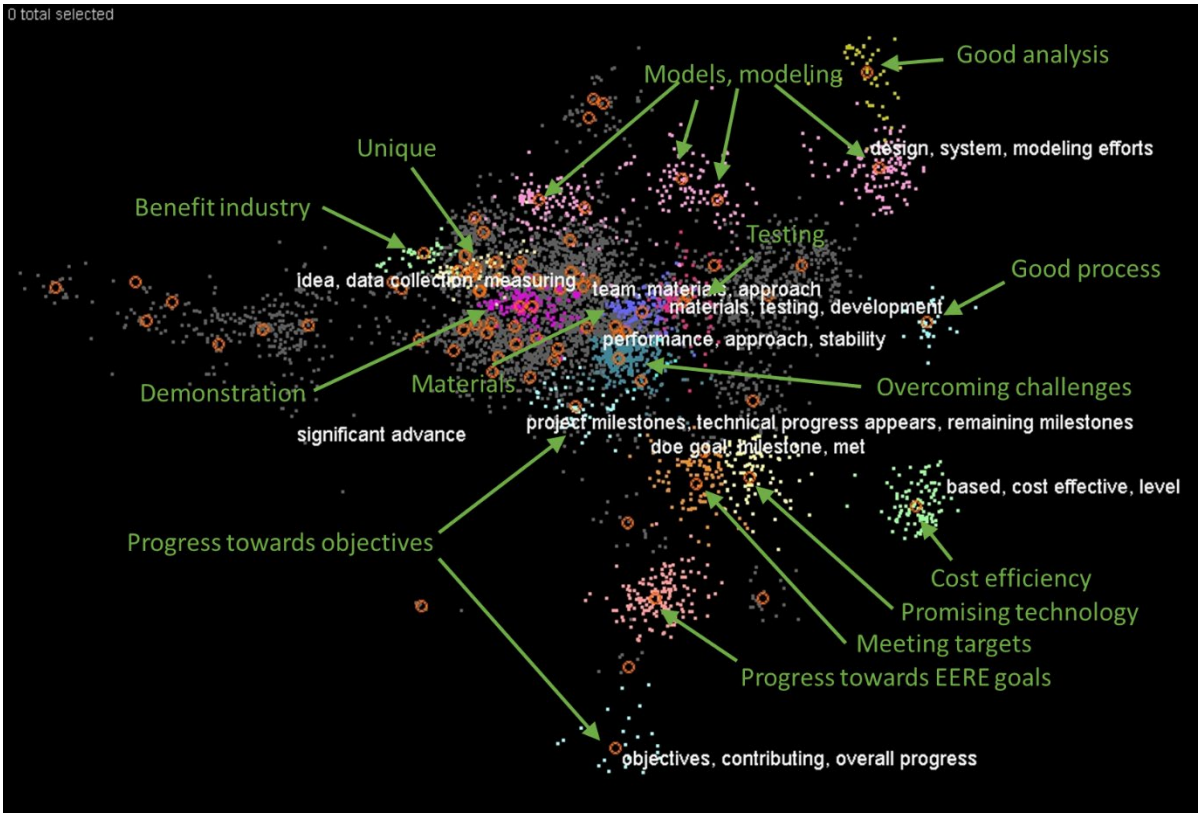

Figure 2. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the strengths of accomplishments.

The following are excerpts from specific reviewer comments taken from the records in each cluster. Percentages shown below are relative to the 6,633 records that contained a strength attribute for a project's accomplishments. Note that approximately 2,200 records (29%) in this dataset contained a strength keyword, but the clusters for these records could not be mapped to an identifiable strength theme.

5.9% of the records contained strengths that mentioned modeling and/or model improvements:

- “This was a good approach to get more variability in the model”
- “Model compounds are useful for identifying promising pathways and doing preliminary process variable scoping”
- “The low-order reactor model capturing multiphase residence time distributions is an impressive accomplishment”
- “Model compound work with button cell effectively demonstrated the proof of concept”

5.7% of the records contained strengths associated with projects facing various challenges:

- “Given the thermodynamic challenges being faced head on by this project, the progress is impressive”
- “As noted, there have been significant challenges and they are being overcome as they come up”
- “Technical progress also appears to be on-track and impressive considering the technical challenges”
- “Progress has been good considering the practical equipment failure challenges”

2.8% of the records contained strengths associated with projects that advanced/made progress towards broader EERE goals:

- “Very good progress toward EERE goals”
- “Overall the team has made good progress towards project and EERE goals”
- “Making good progress toward overall project and EERE goals”
- “Good progress towards all project and EERE goals was made in this reporting period”

2.1% of the records contained strengths associated with projects whose accomplishments are meeting targets:

- “Progress seems to be good: hitting all targets so far”
- “The project met its goal of generating the data needed for evaluation”
- “Overall, excellent work and the project is well on the way to reaching technical targets”
- “The systems have improved significantly and are meeting or exceeding some of the targets at some conditions”

2% of the records contained strengths associated with projects that have a demonstration or have demonstrated something:

- “2000+ hour demonstration of integrated process is impressive”
- “Electromagnetic bench scale test done, successful demonstration of removal of organics”
- “This project is conceptually sound and a good demonstration of fundamental oxygenate conversion chemistry”
- “The most important results were the demonstration the chemistry of the ketonization”

2% of the records contained strengths for projects related to materials and/or material development:

- “Good progress on materials identification/evaluation”
- “The use of composite materials to stabilize the coating was another significant advance”
- “Success in identifying metallic and non-metallic materials resistant to bio-oils is imperative”
- “Good progress has been made on materials, preliminary reactor design, and process modeling fronts”

1.8% of the records contained strengths for projects that are improving cost efficiency:

- “10× reduction in capacity cost is obviously significant”
- “Low cost settling efforts were successful”
- “As the majority of the cost of electrolytic H_2 production is from the electricity cost, targeting system efficiency improvement is a good strategy”
- “If successful the return is large via very significant cost reduction of stack cost and the possibility to switch to cheaper material sets”

1.4% of the records contained strengths for projects whose accomplishments show a promising technology or demonstrate the potential of what can be done:

- “Initial progress has been very good and confirms the great potential of the project”
- “Through-plane reactive excitation efforts are showing good progress and potential utility”
- “The results with the [ ] catalysts were impressive and promising, particularly for [ ] conversion”
- “Good directional results showing the potential for the process”

1.3% of the records contained strengths for projects that demonstrated progress towards a project’s objectives:

- “Previous work on process optimization with dynamics produced good results, which can now be used to achieve the project objectives”
- “The progress made in meeting project objectives and schedule is impressive”
- “Based on their results they made excellent progress meeting the project objectives”
- “Accomplishments and progress by this project towards its objectives have been excellent”

1.2% of the records contained strengths for projects that include testing and recognize the importance of testing activities:

- “Good start on establishing a baseline for their future tests and developing the testing equipment”
- “The accomplishment of testing the process with a wide variety of algal species is on target with the validation of this systems use at scale”
- “Good progress has been made in qualifying the [ ] for testing and in expanding the model of [ ] and solar field performance”
- “Good progress was achieved on assembly and testing of three prototypes, enhanced puck design and fabrication”

0.9% of the records contained strengths that highlighted accomplishments that are unique:

- “This is unique and good”
- “Recycle of materials at pressure and re-liquify strategy is very unique”
- “The machine learning and kinetic modeling adds a unique angle to this project”
- “The systems work is excellent and unique in the project”

0.9% of the records contained strengths highlighting accomplishments with industry benefits, benefiting industrial partners:

- “Project team has been quite successful enabling industry partners”
- “Successfully delivered multiple projects to industry partners, reducing core funding needs”
- “Significant result that industry funding commercialization project”
- “This kind of broad survey is very useful to industry in taking a wide scope”

0.8% of the records contained strengths for projects that have shown good analysis, and useful applicability of the results of the analysis:

- “Pine analysis detail excellent”
- “Analysis of solid and liquid products was very good and provided key insights into the fate of the extractives”
- “I commend this systems-level analysis - actually useful in the real world!”
- “There is a good analysis of the critical issues and ways to overcome them are thought through”

0.6% of the records contained strengths for accomplishments that are demonstrating and developing effective processes:

- “The process is effective in freshwater and even more effective at high salinities”
- “The more feedstock flexible the process the more impressive it will be”
- “Process consistency was demonstrated on the porosity for good reproducibility”
- “A more than 13 times increase in scale is encouraging and good strides towards optimizing the process”

The above examples of specific reviewer comments represent 29% of the 6,633 records. Many of the remainder of the records contained a strength keyword but could not be aligned with a strength theme. Examples include:

- Pretty good so far
- The reported progress looks good
- Excellent, better than intended
- Developed membranes successfully
- Technical accomplishments are impressive
- It has been demonstrated very well with successful results

4.2 Detailed Results of Approach Strength Themes

This section contains the detailed results of IN-SPIRE's analysis of strength themes contained in Approach comments. This dataset contained 6,554 individual records, extracted from 4,663 reviewer comments for 1,359 different projects. Statistically, IN-SPIRE™ organized these records into 80 clusters, which distilled to 14 interpretable themes. Figure 3 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

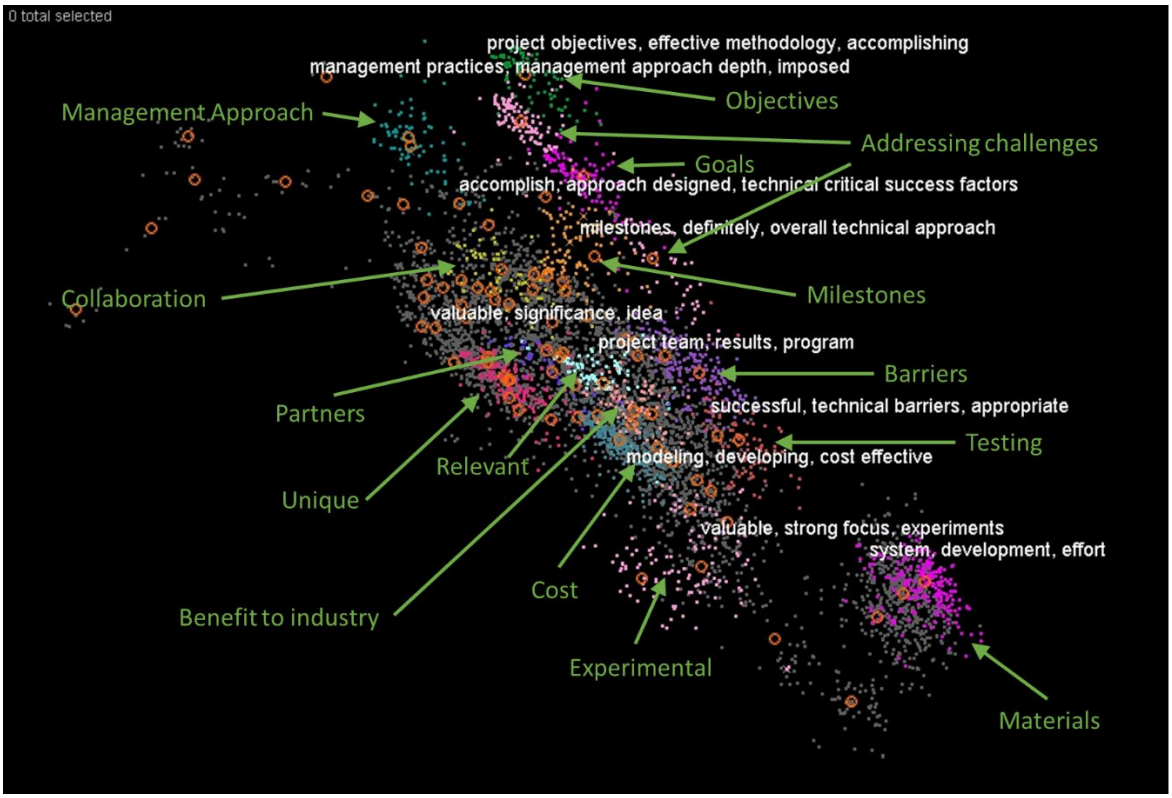

Figure 3. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the strengths of a project approach.

The following are excerpts from specific reviewer comments taken from the records in each cluster. Percentages shown below are relative to the 6,554 records that contained a strength attribute for a project's approach. Note that approximately 4,000 records (61%) in this dataset contained a strength keyword, but the clusters for these records could not be mapped to an identifiable strength theme.

3.8% of the records contained strengths highlighting a cost- effective approach:

- "The technical approach is a cost-effective"
- "It could significantly lower the cost of upgrading of pyrolysis liquids"

- “Project activities nicely linked to potential cost reductions and beneficial impact on program goals”
- “This program has an effective approach to reduce hydrogen production cost”

3.5% of the records contained strengths for an approach that uses or develops materials well:

- “The effort in materials development is outstanding”
- “Generating and validating the models based on an existing material that is near the desired properties is a good approach”
- “The overall approach taken by the project team is excellent, combining material exploration with practical engineering solutions for onboard materials-based hydrogen storage systems”
- “The technique appears to be quite useful in the development of new materials”

3% of the records contained strengths for a project approach that will identify and/or address barriers:

- “The barriers to success are clear described and possible ways to mitigate these have been identified”
- “It would be critical to identify and quantify critical success factors early-on, while progressing on addressing the technical barriers”
- “Identified some key barriers to success including feedstock pretreatment and is addressing this issue”
- “Approach is effective at the laboratory scale and contributes to overcoming most barriers”

2.9% of the records contained strengths for an approach that includes the identification of challenges:

- “Potential challenges and critical success factors are identified”
- “The challenges and success factors identified are reasonable”
- “The team seems to have a good grasp on these challenges”
- “Challenges and success factors are well understood”

2.8% of the records contained strengths for projects with a unique approach:

- “This approach potentially offers a unique pathway for development and deployment of a modular system for methane conversion”
- “This specific effort utilizes a unique technology, anionic resin sheets, distinguishing it from other efforts”
- “Project overview objectives are clear and unique”
- “This project is unique compared to the entire feedstocks and logistics portfolio”

1.8% of the records contained strengths for an approach that includes good experimental design:

- “The integration of modelling and experiment is also to be welcomed”
- “The project includes both experimental and theoretical tasks”

- “The project has strong experimental design guiding research throughout the tasks”
- “The approach is generally good by including both experimental and computational effort”

1.7% of the records contained strengths for an approach that includes effective goal setting:

- “Competence of the team is good, and goals are reasonably ambitious enough”
- “Goals and the audience you need/want to reach is excellent”
- “The project's goals and critical success factors are clearly described”
- “Project goals were clearly described”

1.7% of the records contained strengths for an approach that includes testing:

- “Approach is a good mix of analysis, model testing and real world testing”
- “Good science-based approach starting with virtual test environment”
- “The approach of the medium and heavy-duty field-testing project has proved to be excellent”
- “A good logical and structured approach to real sea testing”

1.6% of the records contained strengths for projects with a good management approach:

- “The team has made a good, realistic assessment of the technical and management challenges”
- “The technical and management approach are well described with milestones and go/no-go's”
- “Management of program follows good program R&D practices”
- “Good project management approach, with clear tasks and milestones for each one”

1.6% of the records contained strengths for an approach that is aligned to meeting project’s objectives:

- “Project objectives were clearly defined with quantitative targets”
- “The project objectives and management approach were clearly defined”
- “Significant progress has been made to meet the project objectives”
- “They provided a very good explanation of the background information for setting up the project objectives and relevance”

1.6% of the records contained strengths for an approach that includes well defined and appropriate milestones:

- “This is a well-planned project with a good distribution of tasks and milestones”
- “The proposed approach is effective and well-articulated, with clear partition into tasks and precise schedule and milestones”
- “The approach identified in the phase 1 and 2 milestones of the project is very good”
- “Progress towards the milestones is good”

1.4% of the records contained strengths for an approach that is generating relevant results:

- “The work is very interesting and relevant. This will help to support our company to produce products that meet not only current portfolio for today, but also a balanced portfolio for the future”
- “A complex task was approached with a focus on imparting maximum relevance to industry and general users”
- “The approach is solid and will likely yield useful results that will be applicable to physical problems”
- “The approach is generally good and is now showing results”

1.4% of the records contained strengths for an approach well aligned to provide benefits to industry:

- “This project is very relevant to benefit the hydrogen economy”
- “The project will help the [ ] program and the fuel cell industry”
- “Focus on developing and deploying technology that will benefits entire industry”
- “The inclusion of real-world tire manufacturing and testing is a critical part of the approach that is essential for moving this work from the lab to industry so its benefits can be realized”

0.9% of the records contained strengths for an approach that leverages strong and successful collaboration:

- “Very strong project structure assuring extensive collaboration of U.S. private and public sectors with strong leadership from private sector with strong collaboration with other energy sectors”
- “It is noted that a lot of effort has gone into collaborating and coordinating between the various national laboratories, and these efforts and results are commended”
- “The technical and management approaches are sound with good skills and expertise in place and great collaboration between the researchers”
- “The project is a good example of academia-industry collaboration”
- “The team appears to have a good plan and a great collaboration in place. The industrial partner is heavily involved which can only help”

0.6% of the records contained strengths describing an approach that makes good use of partners to fill crucial roles within the project, or is well connected to ensure project success:

- “Strong partners plus good connections with large company manufacturer and end-users”
- “Having partners that can fabricate and test actual modules is a significant plus”
- “The team appears to be very strong and includes a combination of academic and industrial partners to cover all aspects of the scope”
- “The management approach involves a strong group of industrial, academic and national lab research partners covering both product manufacturing and commercial commodity issues”
- “This project successfully pulls together partners and consumers for experiences, education, and analysis through surveys”

Many of the remainder of the records in this dataset generally fell into two categories. Some records contained a strength keyword but could not be aligned with a strength theme. Examples include:

- “Overall, an excellent, thought-provoking project”
- “The modules built to fit in shipping containers is good”
- “The plan/approach is good”
- “The approach is good and well thought out”
- “This project is very promising in terms of scope and significance”
- “Many applications would benefit from this approach”
- “Interesting and very thorough approach to determine what influences the process and to quantify its influence on strength and corrosion”

Other examples included a strength keyword, but the benefit was very specific to a particular technology, which made it challenging to find clusters of similarly themed comments:

- “Using [ ] biomass is an additional benefit, since the biomass is generated from water clean up”
- “[ ] certainly have the promise of being superior in this regard to inorganic particles like zeolites”
- “The production of fermentable sugars alongside other useful products is also a potentially significant step forward”
- “The concept of [ ] is an interesting one, with the potential for significant impact on energy use”
- “The project rightly focuses on aerodynamic drag of class 7-8 tractor trailers, which is a significant contributor to fuel consumption”

4.3 Detailed Results of Collaboration and Coordination Strength Themes

This section contains the detailed results of IN-SPIRE's analysis of strength themes contained in Collaboration and Coordination comments. This dataset contained 4,713 individual records, extracted from 3,589 reviewer comments for 1,058 different projects. Statistically, IN-SPIRE™ organized these records into 68 clusters, which distilled to 13 interpretable themes. Figure 4 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

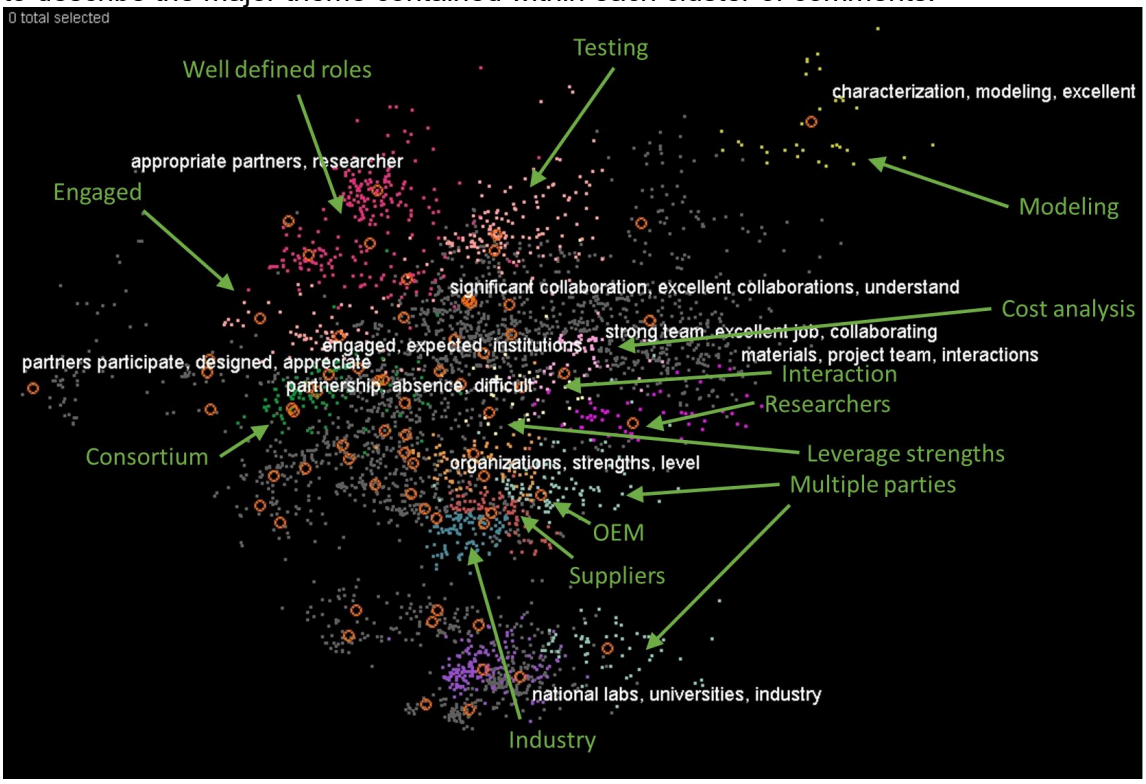

Figure 4. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the strengths of collaboration.

The following are excerpts from specific reviewer comments taken from the records in each cluster. Percentages shown below are relative to the 4,713 records that contained a strength attribute for a project's collaboration and coordination. Note that approximately 3,000 records (63%) in this dataset contained a strength keyword, but the clusters for these records could not be mapped to an identifiable strength theme.

5.8% of the records contained a strength describing well-defined roles for each collaborator:

- "Good collaboration with each party's responsibility well-defined"
- "The collaborators appear to be well-coordinated, and the presenter made clear their respective roles"

- “The responsibilities of the team members are well-defined and the team seems to be working well together”
- “The collaboration between the team members appear to be well-coordinated with each other’s roles clearly identified and focused”

4.5% of the records contained a strength for projects with a good mixture of labs, academia, industry:

- “Good collaboration exists between university, national laboratory, and industry”
- “Well-coordinated effort between [ ] National Lab, [ ] University and [ ] University”
- “This is a good multi-intuitional project involving a university, two DOE laboratories and an industrial partner”
- “The fact that this project involved contributions from 9 national laboratories and 13 universities is testament to the excellent level of collaboration and coordination between the various partners”

3.8% of the records contained a strength highlighting projects with strong industry collaborations:

- “Good outreach over the years to the biofuels industry, and have made critically important contributions to shaping and improving the biodiesel industry”
- “Collaborations with industry are the strength of this project. The various industry partners are highly representative of the significant and active participants in the advanced powertrain industry”
- “Excellent collaboration with the national laboratories and industry. Also leveraging two industry experts to add additional credibility”

3.2% of the records contained a strength for projects that include a collaborator to perform testing:

- “Electrospinning work is progressing, and testing at [ ] has been progressing and is valuable to the project”
- “The collaboration between synthesis and testing centers seems to be working well”
- “The project has a variety of collaborators involved with the supply chain and to assist in testing”
- “The project has excellent collaborators for electrochemical fabrication and testing”

3% of the records contained a strength for projects that included collaboration with suppliers:

- “The team has done a good job at interacting with industry suppliers and has obtained a good variety of materials”
- “They have engaged materials companies and component suppliers”
- “Very good collaborations with part/component suppliers/partners”

2.8% of the records contained a strength for projects with strong ties to a consortium:

- “The list of institutions involved in the consortium is impressive and given the fact they are all connected to each other, it is expected to foster many new collaborative activities that otherwise won't have been possible”
- “Making good use of consortium resources”
- “There is excellent communication and collaboration with many principal researchers in the [ ] consortium, and across all of the [ ] consortium”
- “Existing collaborations with [ ] consortium are good”

1.9% of the records contained a strength for projects with engaged collaborators:

- “The partners provide an appropriate range of expertise and appear to be engaged and participating”
- “There are a lot of team members, including some no-cost participants, and all seem to be engaged”
- “The partners are well-engaged”
- “Good to have the main (current) stakeholders involved and engaged”

1.9% of the records contained a strength for projects that are able to leverage strengths and expertise of each collaborator:

- “Tasks are segmented to appropriately leverage the strengths of each contributing institution”
- “The team will utilize their best strengths and support each other”
- “Contributors are being leveraged well to play to their strengths”
- “The collaboration combines several institutes and groups and is well-coordinated based on the strengths and capabilities of each partner”

1.7% of the records contained a strength for projects that include collaboration on performing cost analysis:

- “The monthly call with team members, [ ] and [ ] National Laboratory was very useful in helping with their cost analysis”
- “The approach taken to work independently on the cost analysis and afterwards comparing the results is very good and persuasive”
- “The project has very good partners, in terms of experience in the industry, technical knowledge of fuel cell systems and manufacturing, cost analysis, and market analysis”
- “The team has set up effective collaborations on cost analysis, film depositions, microstructural characterizations, and performance tests”

1.7% of the records contained a strength for projects that included collaboration with OEM organizations:

- “Project has good collaborators with a material supplier and an automotive OEM”
- “Project coordination with a key stakeholder OEM and a university providing novel nano-fiber support structures is excellent”
- “The team has a good balance of technical expertise from OEM and consultants”

1.6% of the records contained a strength for projects that coordinated with external researchers:

- “There are good collaborations with several researchers within DOE and elsewhere”
- “There are good collaborations with several researchers from the [ ] National Laboratory and also with external researchers”
- “Because this work is connected with other researchers at [ ] University and [ ] Laboratory it merits a satisfactory grade for collaboration and coordination”
- “There are good collaborations with several researchers within DOE laboratories and with universities”

1.1% of the records contained a strength for projects showing with good interaction among team members:

- “There is evidence of strong collaboration and interaction between the project”
- “The team has shown good interaction and expected to handle the project well”
- “There were excellent interactions within the project amongst the partners”
- “Good interaction and collaboration between partners are evident”

0.9% of the records contained a strength for projects that include a collaborator to perform characterization and modeling activities:

- “Excellent collaborations with some groups that utilizes what these groups are good at, such as modeling and material characterization”
- “Good coordination & collaboration between labs, as well as among experimentalists, characterization & modeling”
- “Team has excellent participants well-known in [ ] materials design, electrodes fabrication and characterization”
- “Team has excellent participants well-known in the polymer design, characterization and modeling of fuel cells”
- “The collaboration with the characterization groups is excellent”

Many of the remainder of the documents contained a strength keyword but could not be aligned with a strength theme. Examples include:

- “Collaboration with [ ] is good”
- “Excellent and extensive. All bases are covered.”
- “Collaborations are ongoing and appear to be effective”
- “The current collaboration and coordination is impressive”
- “This project has had outstanding collaboration”

4.4 Detailed Results of Relevance Strength Themes

This section contains the detailed results of IN-SPIRE's analysis of strength themes contained in Relevance comments. This dataset contained 3,281 individual records. Statistically, IN-SPIRE™ organized these records into 57 clusters, which distilled to 8 interpretable themes. Figure 5 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

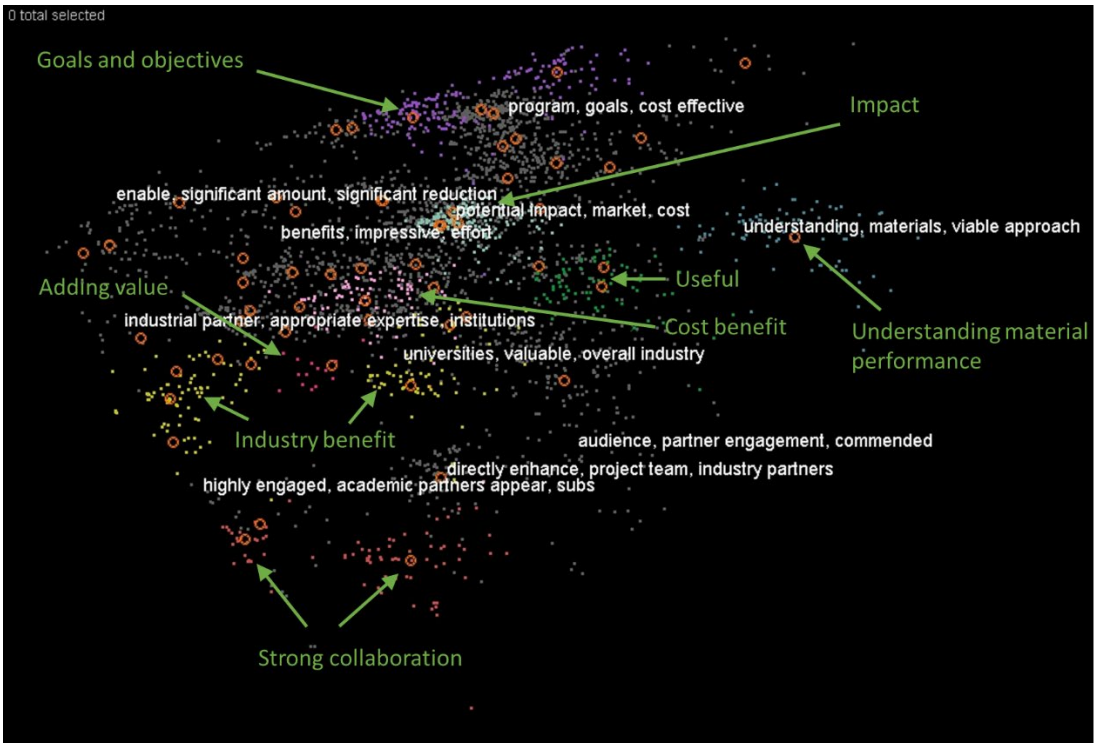

Figure 5. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the strengths of a project relevance.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 3,281 records that contained a strength attribute for project's relevance:

6.9% of the records contained a strength that the project is relevant in meeting larger goals and objectives:

- "This project is highly relevant to the objectives of the [ ] institute"
- "The objectives of this project are highly relevant to goals in the program's Multi-year Program Plan"
- "The project is relevant to the objectives of [ ] and if successfully developed,"
- "Would make a positive impact"
- "This project is highly relevant to the program goals and objectives"

5.8% of the records contained a strength that the project is relevant because of its potential impact:

- “This project is very relevant and will have a large impact”
- “The project is a relevant one and will have a major impact if successful”
- “This is a highly relevant project with, potentially, the largest impact among all ongoing efforts”
- “The project addresses a, potentially, large-scale application, so the impact can be significant”

5.8% of the records contained a strength that the project is relevant for an industry and will have industry benefits:

- “There is very good relevance to the institute, [ ] , in alignment with Multi-year Program Plan goals, and having relevance for the overall industry”
- “The extension of the analysis into machine learning and advanced active process control is the second benefit that the project can provide to industry”
- “Finding a solution to this problem would benefit the industry”
- “This project has made strong contributions to method standardization across the algal industry”

3.0% of the records contained a strength that the project is relevant through its contributions in understanding material performance*:

- “High performance material formulation would have high hydrogen production capacity, meaning high numbers of oxygen vacancies per mole of active material along with fast redox kinetics”
- “The method regenerates the storage material off-board the approach is still highly relevant and efficient”
- “If the investigators continue to integrate current material advancements and fundamental understanding in their model, it will remain highly relevant”
- “The use of an easily applied RECYCLED carbon fiber material system to achieve this end is also attractive”

2.9% of the records contained a strength that the project is relevant due to the team having the right partners, and/or strong collaboration among the partners*:

- “Excellent collaboration with two companies and one national laboratory”
- “The collaborations are strong and represent many different stakeholders”
- “The project has shown excellent coordination with a variety of partners”
- “Given the nature of the project, outstanding collaboration between the two labs, the utilities, the universities, and [ ] Resources Board”

2.6% of the records contained a strength that the project is relevant since it reduces costs*:

* While this is a strength related theme found in relevance comments, it seems less associated with the relevance criterion and more with a different criterion. Future work could align these themes with a more appropriate criterion

- “Reducing battery manufacturing cost would significantly benefit adoption of battery powered vehicles”
- “The focus they are putting on improving the installation, operations and maintenance (IO&M) costs will benefit the entire community. These costs are not always given the importance they deserve.”
- “Increased cylinder pressures and reduced cost for engines and engine materials do support the DOE petroleum reduction goals”
- “This project not only gives novel products for high energy low cost and stable cathode materials of lithium-ion batteries”

2.1% of the records contained the strength that the project is relevant because it is seen as having pre-commercialization and commercialization use:

- “A successful result will provide a tool useful for PI as well as other things”
- “These results are also useful to those in the [ ] conversion technology industry”
- “This website seems to be very useful and is currently being used by private industry and is directly relevant to the Multi-year Program Plan”
- “A useful tool for understanding how to build out some of the industry”

0.5% of the records contained the strength that the project is relevant because it brings value to a process, field or industry:

- “It does add significant value and appears to have potential for some broader application if successful here”
- “Project managers clearly identify and communicate the business case and value proposition”
- “The team has considered the value of the project from several angles, and potential benefits over conventional technology that could favor adoption in the marketplace”
- “If the investigators are successful in developing a much more rapid functional genomics pipeline, that could be of value to the entire field”

4.5 Detailed Results of Future Research Strength Themes

This section contains the detailed results of IN-SPIRE's analysis of strength themes contained in Future Research comments. This dataset contained 3,296 individual records, extracted from 2,781 reviewer comments for 1,162 different projects. Statistically, IN-SPIRE™ organized these records into 57 clusters, which distilled to 8 interpretable themes. Figure 6 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

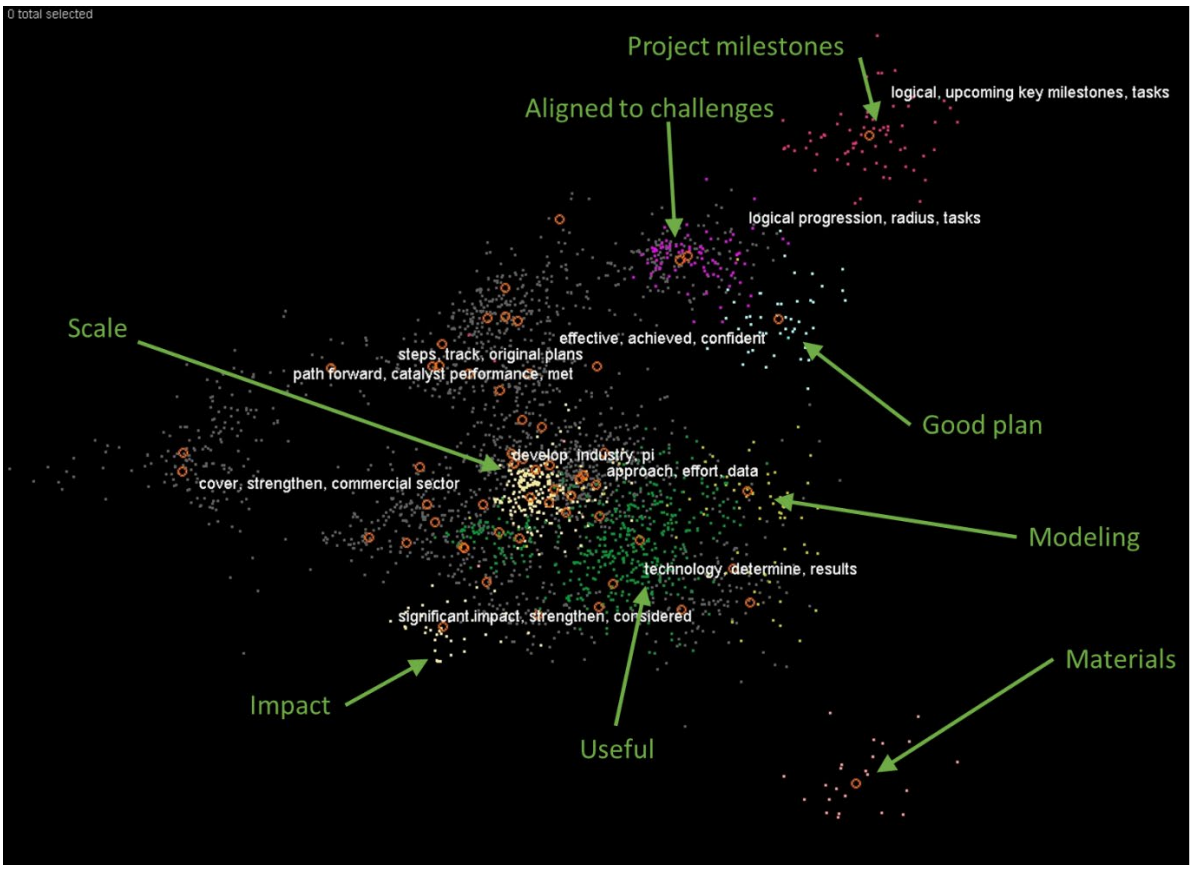

Figure 6. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the strengths of a project's future research plan.

The following are excerpts from specific reviewer comments taken from the records in each cluster. Percentages shown below are relative to the 3,296 records that contained a strength attribute for a project's future research plan. Note that approximately 1,800 records (54%) in this dataset contained a strength keyword, but the clusters for these records could not be mapped to an identifiable strength theme.

15.7% of the records contained a strength that the future work includes an activity that will provide useful information:

- “Detailed multiomics experiments planned, which will be very useful”
- “More deconstruction tests with the formatted versions would provide useful information”
- “The planned scenario/sensitivity analysis including energy prices and incentives will be very useful”
- “Comparison of several different types of compressors would be quite useful”

8.4% of the records contained a strength that the future work will scale-up and/or increase scale of previous work:

- “Scale up of the red mud catalyst testing is also a strong point of the plan”
- “The next phase of work is aimed at catalyst selection and scale-up”
- “Outstanding, very interested in seeing the next steps to scale and make in excess of 650 pounds (lbs)”
- “Scale-up tests are important and good to see included”

2.8% of the records contained the strength that the future work has clearly defined milestones:

- “Using milestones and go no go decisions moving forward is good”
- “The project identified clear goals with quantitative milestones”
- “Future work for each tier is clearly outlined with tasks and milestones”
- “The next period is very important as a number of important milestones will be reached that will define the scope and significance the project results”

2.5% of the records contained the strength that the future work is aligned to previously identified challenges:

- “Planning looks good, with appropriate stage gates, and a good appreciation of the challenges ahead”
- “Good outline of the challenges and opportunities and realistic timelines”
- “The current challenges and the opportunities to overcome them are well-structured and there is high likelihood of success for this project”
- “Proposed future work is in good alignment with remained challenges and barriers as well as with overall goal of project”

2.4% of the records contained the strength that the future work will continue work in model development and modeling:

- “The tasks are logical and well-thought-out. Building the test microreactor and verifying the kinetics is a good way to validate the model and datahub”
- “The future research of validating the model through physical testing is good”
- “The combined experimental and analytical work promises to form the good basis for the conceptual model being developed”

- “The future work looks good, covering access to model process streams”

1.5% of the records contained the strength that the project has a good plan for future research, in general:

- “The plan forward is very good, the task simplified somewhat by the fact that results to date are very promising”
- “The plan forward looks good in terms of development and testing”
- “A good plan is in place to move forward to establishing the value of the platform”
- “The overall plan for future work is varied and solid and it's expected that there will be good data generated and progress made”

1.2% of the records contained the strength that the future work will have potential positive impact beyond the current accomplishments:

- “The proposed future work has clear targets that would have a strong impact on the development in the field”
- “They have good eye to technology commercialization, which should help ensure the impact after the research is completed”
- “The achievement from even part of the listed tasks would make significant impact”
- “Getting to the go/no-go on the group buy program is the most innovative goal in the project and has potential for significant market impact”

0.9% of the records contained the strength that the future work will continue to develop materials:

- “The materials development part is about as good as it gets”
- “Innovative materials as proposed is a good strategy”
- “The future planning was good in terms of set up for further experimentation for evaluation of materials and techniques”
- “Most of the future work proposed is in basic research. This is good, as it is the state in which [ ] materials are at the moment”

Many of the remainder of the documents contained a strength keyword but could not be aligned with a strength theme. Examples include:

- “Planning: good, thorough, well thought through.”
- “The progress to date is significant, and the future plans are well laid out”
- “The test bed validation duration may be a little optimistic but the overall plan and schedule are good.”
- “The 18-month plan looks reasonable to accomplish”
- “The direction of the future efforts looks good”

5.0 Weakness Themes

IN-SPIRE™ was used to identify themes associated with project weaknesses across the five criteria of Accomplishments, Approach, Collaboration and Coordination, Future Research and Relevance. There were many themes in common across these criteria. The weakness themes are summarized below, with details from each criterion in the following sections.

Across the different criteria, there were many recurring weakness-based themes:

- Projects that are failing to address key issues or overcome key challenges or barriers
- Projects that will have difficulties transitioning to industry, technology transfer, marketability
- Projects that were difficult to evaluate due to missing or incomplete information
- Projects that will have challenges scaling from research projects to industrial scale
- Projects that had aspects that were not seen as flawed or not relevant, such as model development, material development, data and measurements
- Projects that were not well-aligned to EERE programs or goals.

5.1 Detailed Results of Accomplishments Weakness Themes

This section contains the detailed results of IN-SPIRE's analysis of weakness themes contained in Accomplishments comments. This dataset contained 2,470 individual records. Statistically, IN-SPIRE™ organized these records into 49 clusters, which distilled to 13 interpretable themes. Figure 7 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

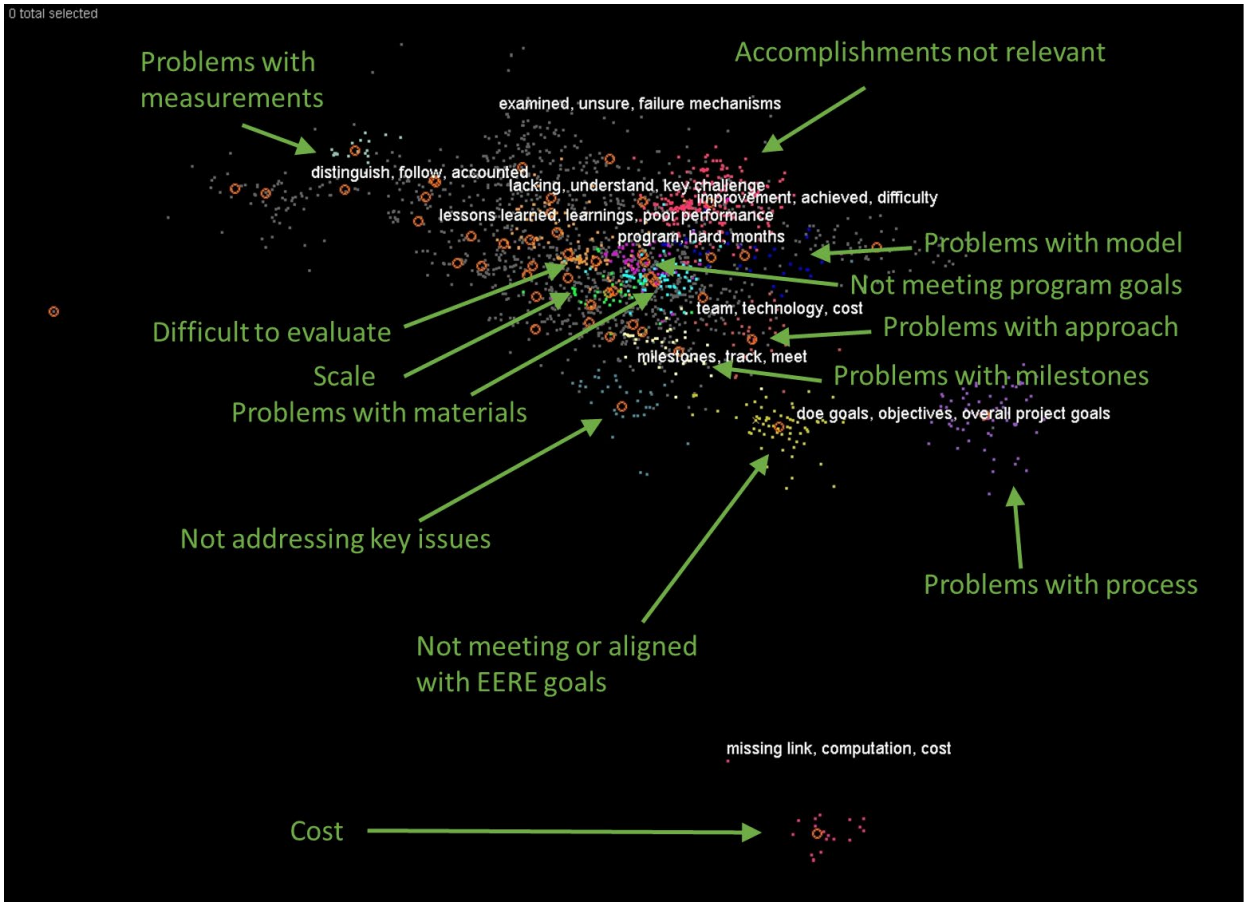

Figure 7. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the weaknesses of a project accomplishments

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 2,470 records that contained a weakness attribute for project's accomplishments:

9.5% of the records contained a weakness indicating the project's accomplishments are not relevant or seen as useful:

- “The completed work might be technically valid, but it is not clear how it is relevant”
- “Not clear from the slide what the objective and relevant accomplishment was”
- “It is not relevant to reduce by 2% when the starting point is near 20% “
- “Work with [ ] was not relevant”
- “I'm concerned that the multi-cylinder work is not quite as relevant for industry”

4.3% of the records contained a weakness indicating problems or struggles by reviewers to assess or evaluate project’s accomplishments:

- “Difficult to judge, as the project is only 3 months old”
- “There are few details of the workplan, so it is difficult to determine the progress”
- “It is difficult to judge the progress for the selected application without any context of a technoeconomic assessment”
- “The results are good, but again, it is difficult to understand them in the way they are presented”
- “It is a bit difficult to judge how some of the older stations reflect what will happen in a retail environment”

3.1% of the records contained a weakness related to a process used by the project:

- “Need for other process operations is not clear”
- “The process is a difficult one, however the low material yield is an indicator of poor progress”
- “The biomass produced does not appear to have a clear link to any particular downstream process or product”
- “It is not clear what the next step in the experimental characterization process is”
- “Electrochemical process seems like it will be difficult to make this work”

3.1% of the records indicated a weakness pertaining to gaps in applicability of materials used or developed by a project:

- “These materials seem to be very poor practical candidates to achieve the EERE H_2 storage targets”
- “It's difficult to fairly compare the results because of the issues with the electrochemical evaluation of the materials”
- “Durability addressed but it's well known that the molecular junctions of dissimilar materials are the weak links”
- “The presentation lacked data on what new courses and materials were being offered”
- “The fact that corrosion is not evaluated early in the project is a significant concern. If it were evaluated, the project could have identified this technical challenge early and the materials design could have been optimized or planned for the future”

2.8% of the records indicated a weakness with a project not meeting goals or its accomplishments were not aligned with EERE goals*:

- “The project failed to achieve a number of its stated goals”
- “It was not clear if the project made any progress against EERE goals this year”
- “They sort of showed a pathway to some of the goals, but it was not clear”
- “Program objectives are being met but it is not clear how these objectives relate to EERE goals”

2.1% of the records indicated a weakness with a project failing to achieve a program goal*:

- “It is not clear how this model control approach can help the program to achieve the program goal”
- “There is no surprise that this program fails to achieve the program goal, because fundamentally the approach suffers from a major flaw”
- “Although the program shows progress with modeling on the engine side, there is lack of specific details with quantitative and measurable results to show the road maps in order to achieve the program goals”
- “It is difficult to see a pathway within this program targeted to achieving durability, since the current trajectory looks like another miss”

1.7% of the records indicated a weakness in defining, evaluating against, or achieving project milestones:

- “The milestones will be difficult to reach”
- “Although there has been some good technical progress here, the overall performance is still very poor. The milestones are really not compelling”
- “This project also suffers from a lack of clear milestones”
- “It is difficult to measure how successful the stated milestones were, such as "Formulate and synthesize redox active oxides from [ ]" without mention of size, properties or performance of the synthesized material”
- “It is difficult to judge achievement of project goals as progress relative to project planning and milestones was not evident”

1.7% of the records indicated a weakness pertaining to challenges or gaps in scale, and/or scaling of a project’s accomplishments:

- “The results do not show much promise and it is not clear how the results would be verified and tested on pilot plant scale”
- “The short time of the scale of the experiments does not allow an accurate assessment of effects on plant operability including equipment-related problems like corrosion”

* These themes appear very similar, but the vocabulary used in the reviewer comments appears to note a distinction between EERE goals and program goals. The comments in the 2.8% theme make very little mention of the term “program”, and the comments in the 2.1% theme make very little mention of the term “EERE”. For this reason, we maintained these as separate themes.

- “It is unclear how these findings will be used to develop a deeper understanding of the system and evaluate the implications and feasibility of scale-up”
- “Focusing on raw biogas from a single source is probably not helpful to scale the process”
- “It is not clear that the reactor design concept is practical in a real-world full-scale situation”

1.6% of the records indicated a weakness related to a project’s model, how it was built, or how it could be used:

- “The incorporation of results into a predictive model for optimized blends is a useful goal, but it is not clear how it was constructed, what the inputs are”
- “It is not clear from the presentation what work has been completed and what is yet to be done with regard to the model”
- “It is not clear how the development of a control strategy or model system for use in vehicles developed by this organization can be reduced to practice by vehicle companies”
- “Your measurements look consistent with failure. “

1.5% of the records indicated a weakness with the overall approach of the project:

- “There was no discussion of how this approach differs from the prior work, or how the challenges of working with an inconsistent, poorly characterized will be overcome”
- “It was not clear why the principle investigator (PI) switched from the previous concept of concrete reinforcement to wire-wrap reinforcement, and the new approach does not look any more feasible to achieve success”
- “It is also difficult to assess progress on their 'foundational' approach, as there is a dearth of quantitative statements as to how far they have 'moved the needle' towards a given target”
- “Although their approach stated that they wanted to first determine what defects and defect magnitudes were relevant, then find methods to detect these, they are in fact operating in the opposite fashion”
- “It was very difficult for me to comment on technical accomplishments as I was having trouble understanding the approach and what the relevance of final numbers that they presented”

1.4% of the records indicated a weakness with a project’s accomplishments failing to address key issues:

- “This feels a bit like throwing a Hail Mary pass. A more prudent approach would be to address and eliminate the manufacturing defects (lack of roundness, poor welds), retest with water, and only then proceed to LH2. Thermal stability and the inevitable boil-off have not been addressed”
- “While the accomplishments in FY17 align with the goals of the project team, they do not adequately address my design or practicality concerns with this vessel”
- “No plans were shown to address low OCV. Performance at lower temperatures is also quite poor and there were no plans presented to address low temperature performance, which would be a big problem during vehicle start-up”
- “One key challenge, which they didn’t really address, is that CV applications are hard to predict—there’s no telling exactly what applications may emerge—as they can include essentially anything that a developer can do with data from the infrastructure or the vehicle”

- “There is also no evidence that the team is doing what should be done to diagnose why the H_2/air performance is so poor, much less that they have a plan to address it”

1% of the records indicated a weakness with cost or cost reporting:

- “It is not clear whether the overall cost increase due to the additional work is simply pushed downstream (and out of the scope of this project) to appear in the final product price, which would not be beneficial”
- “What seemed to be missing, though, is a cost component”
- “What is missing is a cost-benefit analysis of improved yields vs. costs of use of ILs”
- “Missing selected system cost information to support presenting on the cost barrier”
- “Cost of fuel produced is significantly higher than the ultimate price goal”

0.9% of the records indicated a weakness pertaining to issues around the measurements used or generated by a project:

- “Where is the evidence for cavitation in this picture (slide 7)? The PI speaks of 'measurements', though of what is unclear”
- “There are nice pictures of the [ ] reactor to be used for concentrator experiments, but it is concerning that no measurements have yet been done with the cell and/or any measurements under concentration for that matter”
- “Without some more statistical rigor, the validity and reproducibility of these measurements is somewhat questionable”
- “Some progress was made on making stable polymers however, it is not clear how the weight loss measurements is related to fuel cell stability”
- “Some questions arise on the uncertainty evaluation of the measurements. The authors verbally did not seem to be concerned by the uncertainty of the measurements and were unable to explain the translation of the % CO_2 concentration to actual mass of flow estimation”

5.2 Detailed Results of Approach Weakness Themes

This section contains the detailed results of IN-SPIRE's analysis of weakness themes contained in Approach comments. This dataset contained 2,762 individual records. Statistically, IN-SPIRE™ organized these records into 52 clusters, which distilled to 5 interpretable themes. Figure 8 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

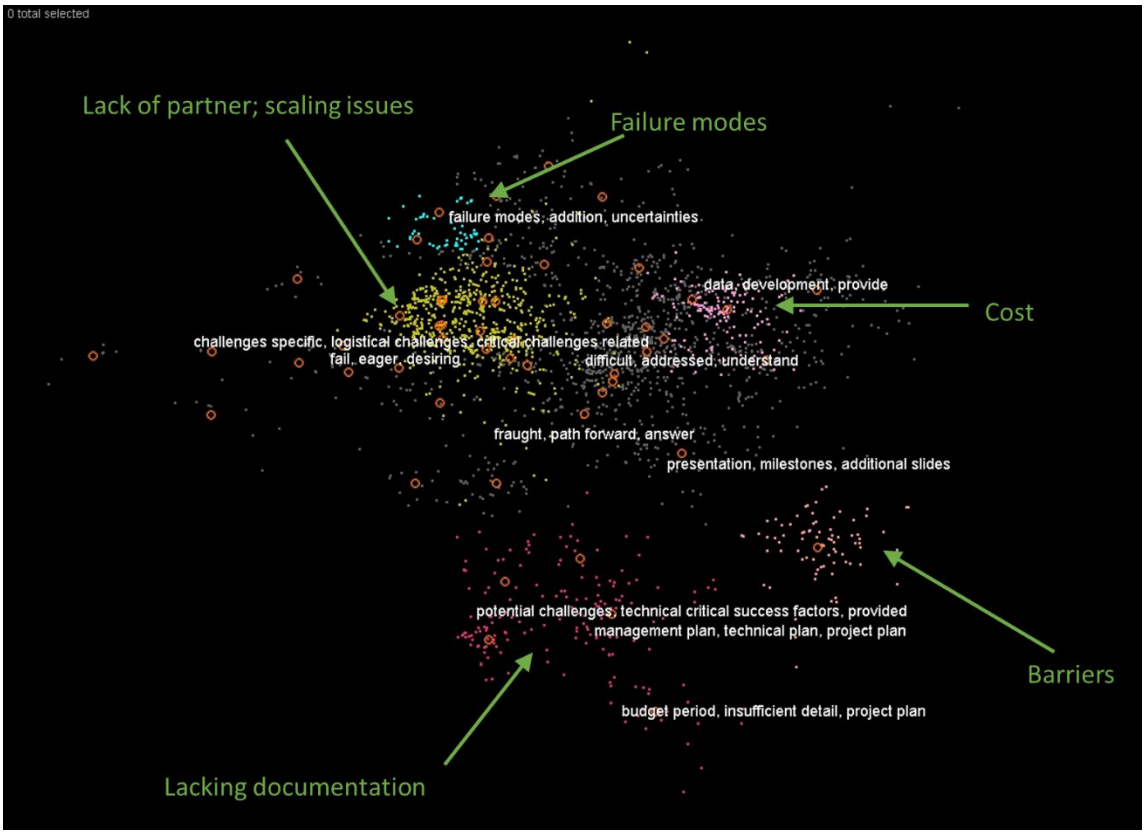

Figure 8. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the weaknesses of a project's approach.

Note: One cluster was not actually a weakness, but instead contained comments about "failure modes". For this cluster, "failure" was not a weakness quality, but we hesitated to remove "failure" from the dataset.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 2,762 records that contained a weakness attribute for project's approach:

30% of the records were related a weakness of either a lack of an industry partner, or a weakness for scaling issues – these themes were intertwined:

- “The lack of planned tests at industry-scale should be reconsidered”
- “It is surprising that the project team did not find an industrial partner”
- “There is a concern that the rare earth metal availability could be an issue in the future as production ramps up”
- “The lack of planned tests at industry-scale should be reconsidered, as it seems premature to assume that bench and pilot-scale behavior is all that is needed for confidence in scaling”
- “This is important, as scale-up has to be a key challenge here, as the experiments so far appear to be only at laboratory scale, and the technology uses ionic liquids, which are not commonly used at large scale”

7.9% of the records were related to a weakness that the project’s approach is lacking documentation of challenges, success factors, project plans, etc.:

- “Top potential challenges were not enumerated “
- “The critical success factors and the top 2-3 potential challenges are not discussed explicitly”
- “Critical success factors and major challenges were not adequately identified”
- “The description lacks detail”
- “A shortfall is that there is insufficient detail about the project plan to judge progress”

5.5% of the records were related to a weakness that the project approach has issues with cost or cost models:

- “A solid cost estimate is needed before it even makes sense to consider taking this approach further”
- “There is a general lack of attention to modelling or including the fundamental elements of cost modelling in the framework”
- “However, it is unclear how the rotating transformer will help meet cost and power density goals”
- “It appears that the cost model for the economic analysis needs to be redone and verified by an independent organization”
- “However, there do not appear to be any specific cost targets identified for the project”

3% of the records were related to a weakness that the project approach is not well aligned towards identifying or overcoming barriers:

- “The approach to improving the material performance and fabrication seems sound but it is unclear how some of their work will affect the barriers in question”
- “It is not clear if these are true barriers to achieving the project objectives”
- “It is difficult to understand technical barriers that were not well communicated”
- “The technical barriers identified are not relevant to the actual work”
- “The technical barriers and challenges have not been discussed and so it the contribution and effectiveness of the proposed approach is vague”

5.3 Detailed Results of Collaboration and Coordination Weakness Themes

This section contains the detailed results of IN-SPIRE's analysis of weakness themes contained in Collaboration and Coordination comments. This dataset contained 736 individual records. Statistically, IN-SPIRE™ organized these records into 26 clusters, which distilled to 4 interpretable themes. Figure 9 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

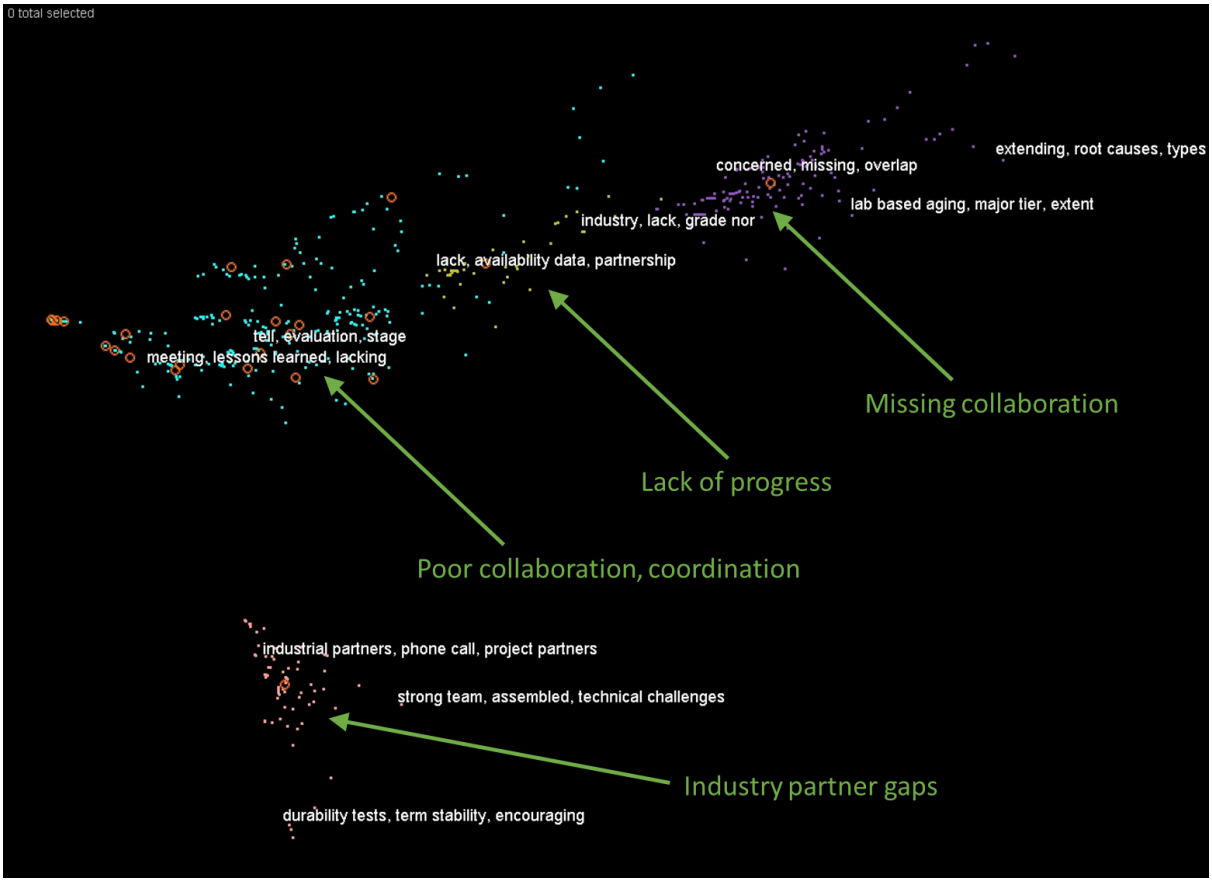

Figure 9. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the weaknesses of a project's collaboration and coordination.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 736 records that contained a weakness attribute for project's collaboration and coordination:

Over 50% of the records were related to a non-specific weakness of the project showing poor collaboration and/or poor coordination:

- "Lack of coordination to gain insight into the many OEM requirements is lacking

- “Collaboration and coordination with partners seem to be lacking or not shown “
- “Collaborations with a potential manufacturer was lacking”
- “However, actual collaboration seems lacking between the different contractors”
- “However, there appears to be lack of collaboration with a partner that can quantify the efficiency and losses”

24% of the records were related to a project has a missing collaboration:

- “A missing ingredient is collaboration with existing capability for detailed characterization”
- “Evidence of collaboration is missing”
- “One expertise missing from the team is a partner who can perform electrode modeling”
- “There remain key collaborators missing to maximize the usefulness of this project”
- “It may be that the team is missing out on the industrial perspective”

10.2% of the records were related to a project having gaps with industry partners:

- “The efforts in reaching out to industrial partners is weak”
- “The technical guidance and commitment from OEM side seems insufficient”
- “Only weakness is perhaps the lack of a [ ] supplier”
- “Operator issues suggest a poor interface or commitment by collaborators”
- “The project has lacked some key vehicle/OEM partners needed for project success”

8.3% of the records were indicating a project’s collaborations or coordination is a source for showing a lack of progress:

- “It seems that there is a lack of progress”
- “Although the number of participants and degree of collaboration is very good, it may also be the source of the lack of progress.”
- “I question, though, the lack of collaborations and coordination with experts in this field”
- “Again, given the lack of project leadership and technical team, collaboration is not significant”
- “The presented results however lack these detailed model outcomes”

5.4 Detailed Results of Relevance Weakness Themes

This section contains the detailed results of IN-SPIRE's analysis of weakness themes contained in Relevance comments. This dataset contained 1,099 individual records. Statistically, IN-SPIRE™ organized these comments into 33 clusters, which distilled to 9 interpretable themes. Figure 10 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

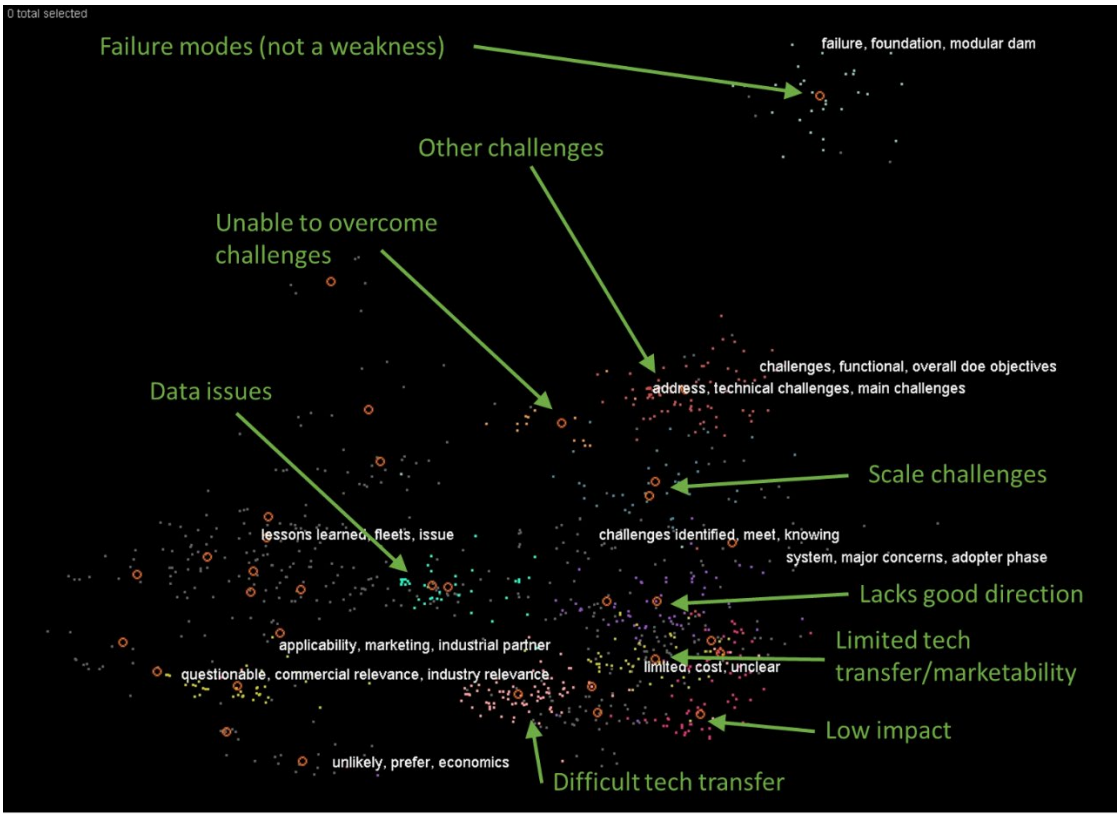

Figure 10. This galaxy visualization from IN-SPIRE™ shows the different clusters of comments related to the weaknesses of a project's relevance.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 1,099 records that contained a weakness attribute for project's relevance:

Note: One cluster was not actually a weakness, but instead contained comments about "failure modes". For this cluster, "failure" was not a weakness quality, but we hesitated to remove the "failure" from the dataset.

9.7% of the records were related to the project having limited technology transfer/marketability potential:

- “The technology transfer/marketability are limited since the industrial partner is not interested in marketing”
- “Technology transfer/marketability will be limited for the application considered”
- “Applications may be limited to only a few systems”
- “The opportunity for collaboration might be limited”
- “Microreactors are notoriously difficult to transfer to market”

9.0% of the records were related to the project having difficulties with technology transfer:

- “The technology transfer is difficult”
- “Technology transfer/marketability will be difficult”
- “I believe this to be a very difficult technology to commercialize”
- “Whether this tech is transferable is difficult to say”
- “Whether this process will be transferable to market is difficult to say”

8.8% of the records were related to the project work not being well aligned to overcoming challenges[*]:

- “There are a number of technical challenges to overcome”
- “Given that the technical challenges can be overcome”
- “Hopefully the project team will be able to overcome the many challenges ahead”
- “It still has many challenges that need to be understood and overcome if it is to successfully make it from the laboratory to the market”
- “The project still has the potential to achieve significant outcomes if it can overcome the current challenges”

6.5% of the records were related to the project having low potential impact:

- “The relevance and the impact of this project is quite limited and will not lead to the targets expected from this institute”
- “Unfortunately, the operational difficulties of the units built for this project limits their impact on the biomass conversion programs”
- “It is unclear of what potential impact [ ] will have commercially”
- “Given the uncertainties associated with the approach indicated above, the potential impact is still rather unclear, however”
- “The impact is limited as the systems out there are already losing very little heat and have low thermal conductivities”

[*] While this is a weakness related theme found in relevance comments, it seems less associated with the relevance criterion and more with a different criterion. Future work could align these themes with a more appropriate criterion

5.2% of the records were related to the project lacking good direction, or having a challenging alignment with advancing industry:

- “My only concern is that market dynamics change quicker than the research direction can to answer a specific question”
- “But, in the absence of a far larger scope that works to understand the fuel cell system, this effort cannot have much impact”
- “It is difficult to see how the thrust area becomes relevant to advancing the fuel cell program”
- “While the project is making good progress towards advancing activity and performance, durability has only been minimally addressed”
- “The project is relevant to DOE's technical barriers, but the approach is quite a longshot and will likely find a difficult adoption path with industry”

3.8% of the records were related to the project having issues with data being used or generated:

- “The lack of preliminary TEA and life cycle data is a concern”
- “I believe the project is a lower technology readiness level than what the authors imply (but not state directly) especially because of lack of clear data showing the scalability potential”
- “It seems that most of the validation is represented by only a single experimental data set and this is not sufficient”
- “However, the study execution lacked detailed data to back the conclusions”
- “Data processing is a big challenge to reduce costs of monitoring”

3.1% of the records were related to the project specifically facing challenges with scaling:

- “There are concerns over how this technology scales”
- “However, it is difficult to foresee a path to low-cost large-scale hydrogen production using this technology that is achievable in a reasonable time frame”
- “There remains significant uncertainty that magnesium sheet is a viable/sustainable material for the high volume manufacture of automotive closures”
- “Challenge in large-scale projects is production of sufficient materials in a consistent manner to provide industry to evaluate in larger-scale tests such as engines and or scale up of other operational units require feed to run as well to produce sufficient fuel for testing”

1.3% of the records were related to the project facing many challenges, in general, which make it difficult to measure its relevance:

- “The technology is marketable if successful, but there are a range of technical challenges here”
- “The challenges appear to be mainly technical”
- “To me this type of process is a long way from market, which makes the prospects for technology transfer difficult to quantify”
- “Several challenges ahead”
- “However, the challenges are equally large”

5.5 Detailed Results of Future Plans Weakness Themes

This section contains the detailed results of IN-SPIRE's analysis of weakness themes contained in Future Plans comments. This dataset contained 1,932 individual records. Statistically, IN-SPIRE™ organized these records into 42 clusters, which distilled to 9 interpretable themes. Figure 11 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

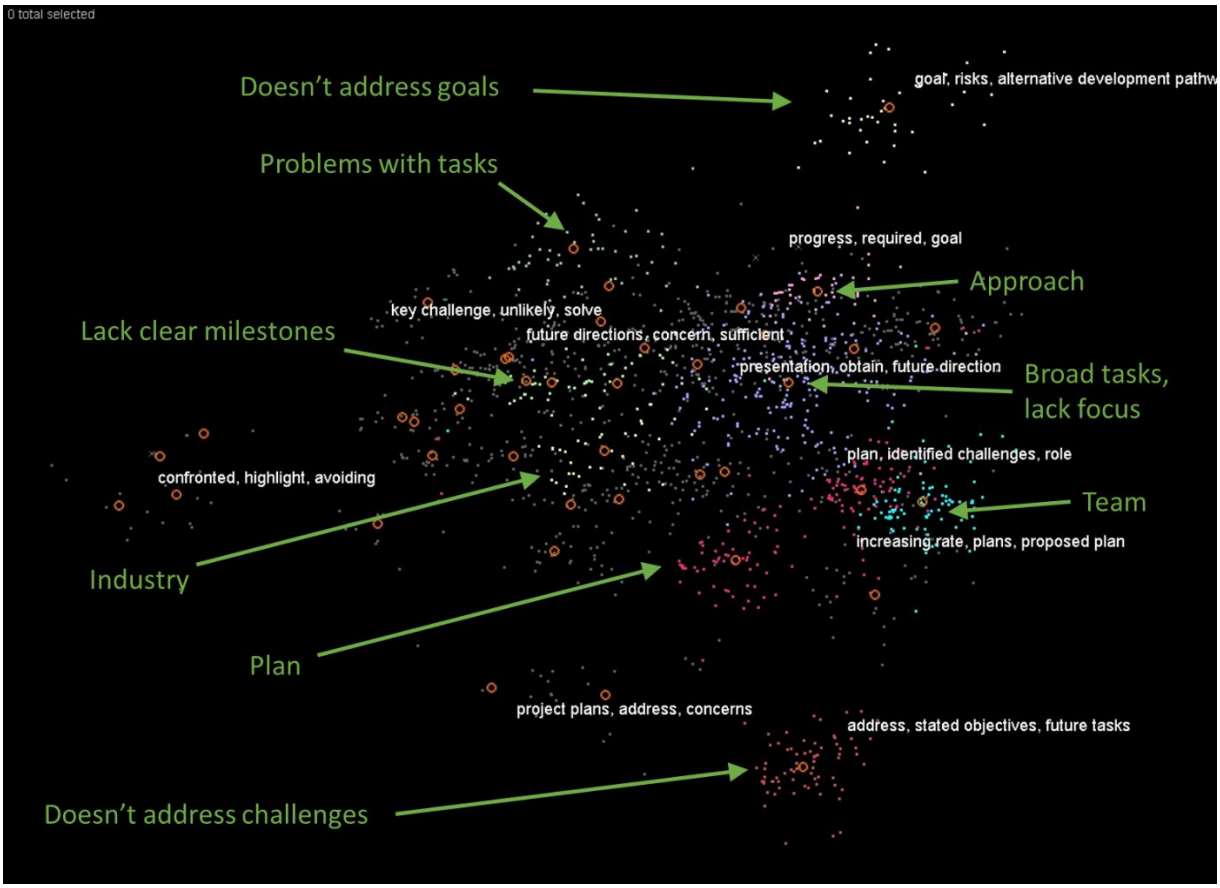

Figure 11. This galaxy visualization from IN-SPIRE™ shows the different clusters of records related to the weaknesses of a project's future plan.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 1,932 records that contained a weakness attribute for project's future plan:

14.6% of the records were related to the future plan tasks being too broad, or that the researchers need to fine-tune their focus, or lack of information:

- “The project has identified a number of tasks that appear logical and may be worthwhile. However, it has not shown how these future tasks fit into an overall strategy or help to reach important goals. Rather, they appear to be just a random set of additional tasks”
- “The future work includes a lot of very difficult tasks; CCM making, degradation mechanisms, catalyst making, modeling, etc. I fear the size and time frame of this project will not allow completion of these tasks. I'm curious what tasks will realistically be completed by the end of the project”
- “Proposed future research focuses on key challenges, but little details of specifics are provided for the future tasks”
- “The proposed research scope is broad and many challenges are pointed out. However, detailed approaches to address these challenges are unclear. It looks to be very challenging to address these open questions. A realistic and achievable future research plan is needed”
- “However, the scope may be too expansive and it may be beneficial to narrow it somewhat”

9.7% of the records were related to difficulties understanding the project’s future plan:

- “The plan as presented in the PowerPoint lacks detail. There are no clear stage gates etc. This needs to be remedied”
- “What is lacking in the future plan is: a) an integrated design solution also showing how a pure water stream is obtained and b) a techno-economic assessment of the integrated design”
- “The future work plan is unclear”
- “The plan for future work was vague, in large part due to missing milestones”
- “The future work plan is weak in my view and not very clear on the final goals”

5.2% of the records were related to the team appearing unprepared to take on aspects of the future plan:

- “The team may be underestimating the challenge of identifying enzymes that work preferentially on long chain compounds”
- “It is unclear how the team is ‘anticipating analyses that are relevant to industry’”
- “However, considering the remaining limited resources and time, it is not clear how the project team plans to accomplish the preparation, submission and respond to peer-reviews of all 13 papers in such a short time”
- “Proposed future work looks good and aligned with project milestones but role of other team members is unclear”
- “It was not clear to me if the team understands the cause of the observed capacity fade”

4.6% of the records were related to the future plan not addressing challenges:

- “It was unclear whether the work on molecular mechanisms will be able to address the transience of phenotypes”
- “The proposed future work makes no sense in light of the current poor performance “
- “There still seems to be a number of challenges remaining for this project to be successful, not certain if they can address everything”

- “No new ideas proposed to address the main challenges. The ideas listed seems more of the previous approach”
- “They need to identify the number one challenge limiting electrochemical CO_2 reduction and say how they will meet it. Their current plan does not address this challenge”

3% of the records were related to missing information and/or other deficiencies in the approach described in a future plan:

- “The present plan misses an integrated approach to the four challenges”
- “Insufficient details, especially on planned approach and risk mitigation are given for future work”
- “Some concerns with the computational approach were noted in an earlier comment”
- “In other words, the mechanism-based approach to understanding degradation is missing”
- “The national level analyses that need to be performed have a large degree of uncertainty “associated with them. It is not clear that the chosen approach can provide the required level of confidence”

2.4% of the records were related to the future plan not addressing goals for program, project, or EERE:

- “Not clear how the enhanced 20% methane uptake goal differs from the 20% enhancement in methane oxidation goal already achieved. Future work goals appear to be sound and will build on results to date, though more details on how these goals will be pursued would have been helpful”
- “The plan needs to be more concrete. It is not clear how the different topics and objectives are related to achieve the overall goals”
- “Future work was not clearly presented in a slide and linked to project goals and milestones”
- “It is not clear what additional work will be done to achieve the stated goal of making the process commercially viable or how the target cost will be achieved by the stated future work”
- “It is not clear how their work so far moves them closer to their monolithic goals, or why they are going through this intermediary step”

2.4% of the records were related to the problems with future plan tasks being addressed or defined:

- “One absolutely critical task is … There is very little point to the rest of the work if a stable chemical-looping material cannot be developed. This task should have been the earliest and highest-priority”
- “Quantitative targets were not identified for each task”
- “It is not clear who is doing each task in the future research steps”
- “It's not clear how this connects with the other goals of this sub-task and the plan to do this was not discussed”

2.4% of the records were related to poor alignment of future plans to industry needs, industry adoption:

- “Unclear how relevant the proposed creation of hardware-and-software-based fallback and contingency operating modes will be to industry and the likelihood of adoption among industry”
- “Some other key industry challenges have not been regarded”
- “It is not clear, how the project includes or addresses potential system operators in the maritime industry”
- “The future work will not lead to any new information within the industry and therefore add no value to addressing barriers and challenges”

1.9% of the record were related to the future plan not containing clear milestones:

- “The future work statement lacks clear milestones and only hints at the reactor-receiver”
- “Additionally, since specific milestones were not included, it was a challenge to discern what the path forward would be”
- “Milestones and decisions points were not clear”
- “There are not clear milestones for success moving forward but each task clearly has logical and beneficial improvement”
- “Lots of future work planned but it is unclear what the go/no-go milestones are or what the metrics for success are”

6.0 Themes Identified for Project Recommendations

IN-SPIRE™ was used to identify themes associated with project recommendations across the five criteria of: Accomplishments, Approach, Collaboration and Coordination, Future Research and Relevance. There were many themes in common across these criteria. The recommendation-based themes are summarized below, with details from each criterion in the following sections.

Across the different criteria, there were many recurring recommendation-based themes:

- Recommendations for increasing the scale of projects to match industrial scale
- Recommendations for more industrial collaborations
- Expanding the use of the data collected
- More comparisons of data to other processes
- General recommendations about the presentations and presentation formats

6.1 Detailed Results of Accomplishment Recommendation Themes

This section contains the detailed results of IN-SPIRE's analysis of recommendation themes contained in Accomplishment comments. This dataset contained 1,222 individual records. Statistically, IN-SPIRE™ organized these comments into 34 clusters, which distilled to 8 interpretable themes. Figure 12 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

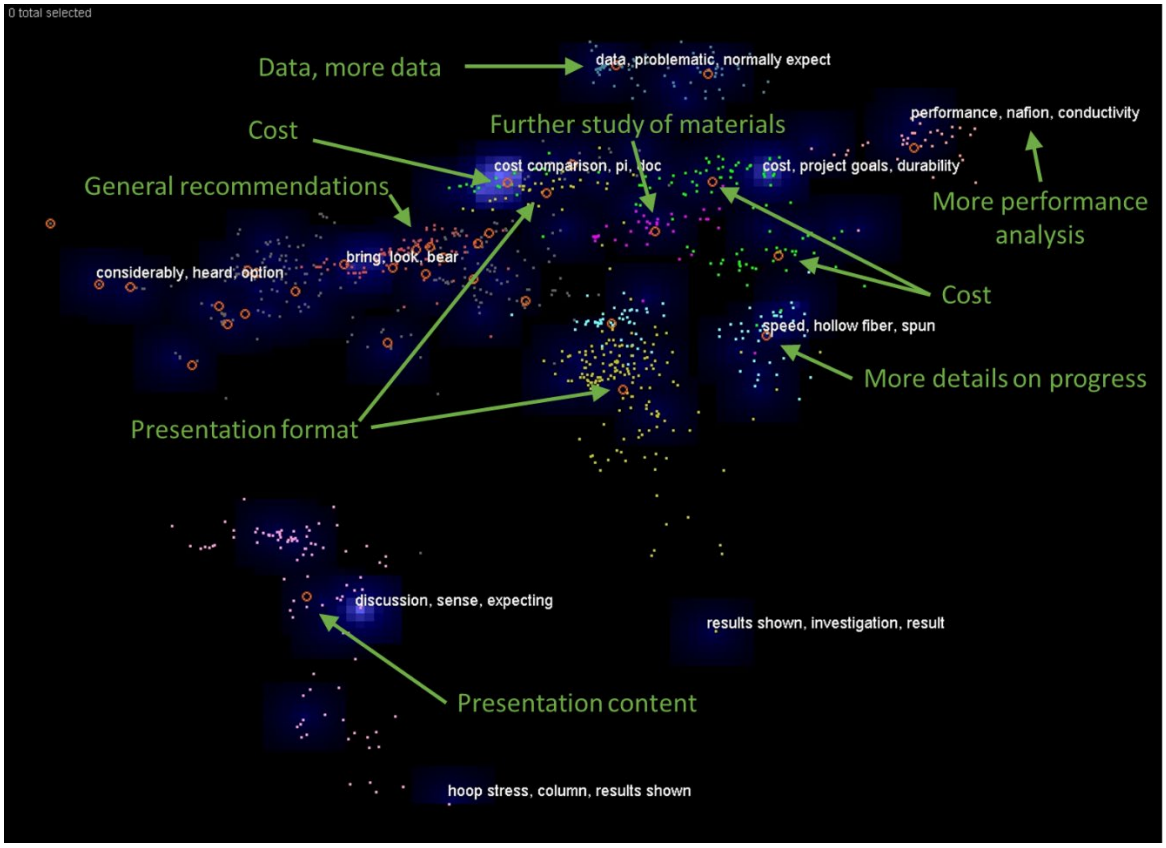

Figure 12. This galaxy visualization from IN-SPIRE™ shows the different clusters of comments containing recommendations related to a project's accomplishments.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 1,222 records that contained a recommendation for project's accomplishments:

17.7% of the records were related to recommendations for changes to the presentation format:

- "Suggest either further breaking up the presentation, elongating the presentation time, or adding more detail to each project in the reviewer only slides"

- “Maybe just the style of the presentation, but I would have expected an emphasis on partnerships, results”
- “It would have been helpful to include comments from previous year’s presentation and actions taken to incorporate their suggestions into this year’s effort”
- “The presentation would have benefitted from a current time line, showing the two budget periods and the go/no-go decision point”
- “You might consider extending the time for the talks to 25 minutes next year, in order to allow time for the presentation and also entertain questions”

11.9% of the records were related to recommendations related to cost, cost assessment, cost comparison:

- “It would be important though to maintain a check on the process times being tested in the enzyme reactions in relation to a cost model”
- “I would also like to see some sort of cost assessment now that they have narrowed down to one approach”
- “I would like to see more sensitivity analysis of electrolyzer capital cost vs. capacity factor”
- “This reviewer suggests that the PI develops cost of model of the process”
- “One thing I would like to see is the cost target for these magnets and whether they can be achieved in near term”

9% of the records were related to recommendations for changes to presentation content:

- “I would like there to be a lot more information on the project plan”
- “It would be nice to have a clear statement of what specific accomplishments were done since project kickoff”
- “The presentation would have been stronger if the integration and collaboration between bench scale results and the pilot results was made clearer”
- “It would be helpful if a matrix was prepared indicating all of the tests that were run, including processes looked at, feedstock blends used, % solids, and operating conditions”
- “Compared to the 2014 presentation, a significant progress in the 2015 presentation is hardly evident”

8.3% of the records were related to recommendations for more details on progress that was made:

- “It would be helpful what progress has been conducted beyond the knowledge of [ ] assessment”
- “I would have like to have seen more progress on completing the design”
- “I would have liked to see more discussion of progress on the gap analysis (Task 3) and a more complete description of the models employed”
- “The reviewer would like to see any progress on accuracy and speed since these are part of the project objectives”

- “It would be interesting to see similar progress on understanding the barriers to high load LTC operation”

7.2% of the records were general recommendations:

- “I recommend they improve the content of their slides for better clarity”
- “I would encourage the PIs to continue with collecting data with the same quickness”
- “It is recommended that they improve their kinetics to decrease the cycle time”
- “I encourage an expansion of the results to include years prior to 2020 so that we can clearly understand the near-term challenges over the next decade”
- “Would like to have seen how the detailed designs and prototyping met expectations”

6.5% of the records were related to data recommendations, include more data, do more stuff with or find more uses for the existing data:

- “I would be interested in seeing that side by side with the bus data”
- “I would have liked to see more preliminary cycle data”
- “I would suggest calculating some binding energies from this data”
- “The only suggestion is to include any cylinder-to-cylinder variation data along with any validation data”
- “This reviewer would suggest that there should be additional emphasis on getting data in front of potential adopters in a more aggressive way”

3.8% of the records were related to recommendations for more performance analysis:

- “It would be good to see fed-batch conditions used early on for the [ ] process to test strain performance, even in a small-scale”
- “Make the performance using [ ] processing as good as it can be, and then compare with the [ ] processing”
- “As the vehicles become ready for deployment it would be good to see more standardized testing across vehicles to enable more direct performance comparisons between them”
- “This reviewer would like to see sensitivity studies on MEAs addressing durability and HCD performance”
- “It is recommended to show experimentally/analytically the presence of graphene and present a model how this increases the performance and/or lifetime”

3.5% of the records were related to recommendations for further study of materials:

- “It would be useful to have a preliminary cost/performance model for the silicon materials”
- “This project would benefit from an understanding of the kinetics involved for the different formulations, so that a model can predict the optimal combination and composition of binary salt material”
- “But it is recommended to think about additional analytical techniques to further strengthen the understanding of the synthesis parameters and link it to the material characteristics and finally electrochemical and lifetime results”

- “The study of solid state synthesis from hydroxide precursors makes sense, but additional information on materials synthesized from carbonates or in solvothermal processes (as the project title suggests) would be equally valuable”
- “The reviewer would like to see such implementation extended to other more promising systems such as other positive and negative electrode materials”

6.2 Detailed Results of Approach Recommendation Themes

This section contains the detailed results of IN-SPIRE's analysis of recommendation themes contained in Approach comments. This dataset contained 1,150 individual records. Statistically, IN-SPIRE™ organized these comments into 33 clusters, which distilled to 9 interpretable themes. Figure 13 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

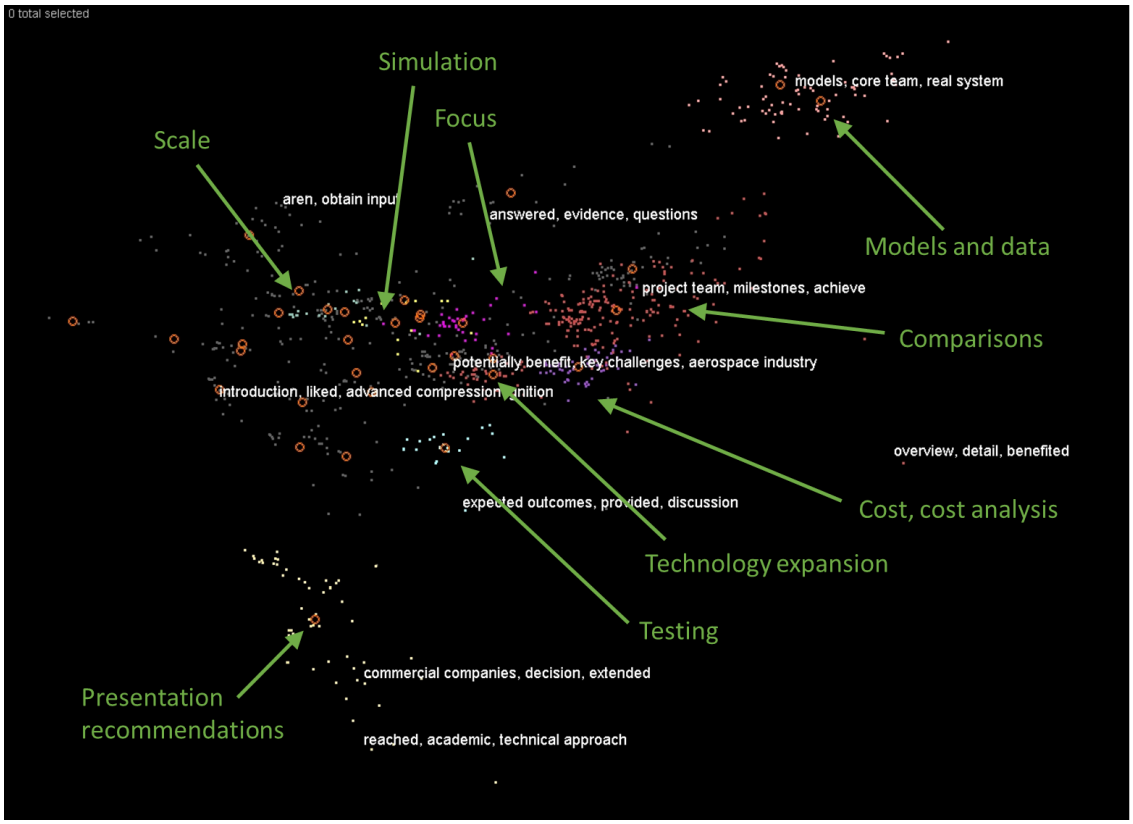

Figure 13. This galaxy visualization from IN-SPIRE™ shows the different clusters of comments containing recommendations related to a project's approach.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 1,150 records that contained a recommendation for project's approach:

22% of the records were related to the project approach needing more information about comparisons to baselines or related technology:

- "It would be helpfully to quantify your results by creating a table (this table could evolve) to show the comparison points to the baseline"

- “I would suggest a comparison with the best (published) membrane based CO_2 capture for the proposed membrane based separation”
- “In the development and application of the diagnostics mentioned, it would be relevant to compare capabilities against competing technologies”
- “I would also like to see some comparison back to conventional ignition systems”
- “Recommend comparing dynamic skip fire vs. the authors intended baseline without cylinder deactivation and then compare to modern engines with fixed cylinder deactivation”

7.5% of the records were related to recommendations associated with models and data:

- “An additional focus is to develop recommendations for extension of current models to include emerging travel technologies and practices”
- “Please create a slide next time that describes the limitations of the models”
- “It would be helpful to provide more information about how the model will be validated, and what data would be used in the validation”
- “If the sole intention is developing a highly accurate predictive model, then it is suggested to perform massive high-throughput experiments to collect a large volume of data for model training”
- “I would like to see a statement that more clearly addresses the expected limitations of the models”

6% of the records were associated with presentation recommendations:

- “Presentation very good, clear and easy to follow. Technical and management approach appear to be solid. Suggest using a Gantt chart to show schedule and task break-down “details.”
- “Some of the test conditions were not apparent in the presentation”
- “I do have concerns regarding the 5% fuel economy increase with the [ ] systems. There wasn't thermodynamic data supporting this number presented in the presentation, and it seems like an aggressive goal. It would be good to know the assumptions that go into the projected 5% fuel economy benefit from this project”
- “I would avoid the use of the term green manufacturing process in a scientific presentation.”
- “But what seems to be missing is a clear vision (or at least presentation thereof) on how all this comes together at the end of the project”

5.5% of the records were associated with needing more information on cost, cost comparisons, or cost analysis:

- “In support of this approach, it would be beneficial if on-going system cost analyses were being conducted to gauge progress toward cost-effectiveness”
- “In addition the cost reduction at GDL level should be huge to impact strongly on the stack cost: the expected decrease could be clarified”
- “It would be beneficial to include some generic cost comparisons, even if not for the materials used here, to demonstrate that the approach used in this project should result in cost savings”
- “I expected to see more of a pursuit towards actual cost-benefit analysis”

- “It would be beneficial to present a cost-of-operation benefit that also considers the cost of the additional hardware”

4.7% of the recommendations were associated with taking a deeper look at technology, or expansion of research about technology:

- “I would encourage the development of a wider set of technologies”
- “Recommend finding benchmarks with existing methods/technology”
- “The technology, if it is to be distributed as suggested will have to be affordable - this needs looking into”
- “Also would be useful to see what technology benefits 2020 has over 2010”
- “The project would benefit by including a technology review of previous work”

3.8% of the records were related to the approach needing to focus in a particular area or direction:

- “I would agree with the researchers that a key challenge is focus - what to focus on and why”
- “Suggest maintaining heavy focus on finding redox materials with improved performance -- don't let this focus slip”
- “I suggest picking one or at most 2 most promising technologies and focus on further developing them and scaling them up to the point that full-scale motors can be built”
- “My recommendation will be to focus the project on 1 or 2 key process steps and understand them well to advance the state-of-art”
- “This work seems diffused with multiple non-overlapping goals - it is recommended to focus on the synergistic catalyst effects”

2.3% of the records were related to the approach needing to include more testing:

- “The project would benefit by testing the reactivity of other pretreatment lignins”
- “If the program were geared toward developing a test standard or methodology to evaluate any flowmeter and included some industry agreed upon acceptance criteria for the testing, it would be of considerably higher value”
- “Suggest an increased emphasis on transient testing, including some several operating regions from certification test cycles”
- “I would like to also see noise transmission testing”
- “I would have been interested to see the effect of testing to a set number of cycles”

2% of the records were related to the approach needing more details with scaling efforts:

- “I would like to see something on the final scale, and associated scale-up”
- “I would heartily recommend integrating this effort across other pilot facilities and where possible with small scale and bench scale as well”
- “It would be good to know what levels the chosen organism can tolerate as a maximum titer for harvesting at scale”

- “The project is one in which it would be important to have a good view of the desired end-state for the production strains, and to get some input from industry (what will be important for economics at scale)”

1.5% of the records were related to the approach needing more simulations, or details about the simulations being performed:

- “It would be very helpful to list out the metrics for evaluation of the simulation and judge the quality/validity of the simulation”
- “It would be advisable to update the simulations and present updated results each year”
- “It is recommended the developed motor model and its parameter are used for some advanced simulation tools such as HIL (hardware in loop) to assess through simulation and see how motor could performs when driven by inverter”
- “It would be good to show the results of the simulation on top of the experimental results that were shown”
- “It is recommended that better consideration of the simulation/experimental cross-validation approach be addressed now to avoid excessive iteration between the modeling and testing”
- “It would be helpful to better quantify time scale effects such as engine speed and injection pressure”

6.3 Detailed Results of Collaboration and Coordination Recommendation Themes

This section contains the detailed results of IN-SPIRE's analysis of recommendation themes contained in Collaboration and Coordination comments. This dataset contained 745 individual records. Statistically, IN-SPIRE™ organized these comments into 27 clusters, which distilled to 6 interpretable themes. Figure 14 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

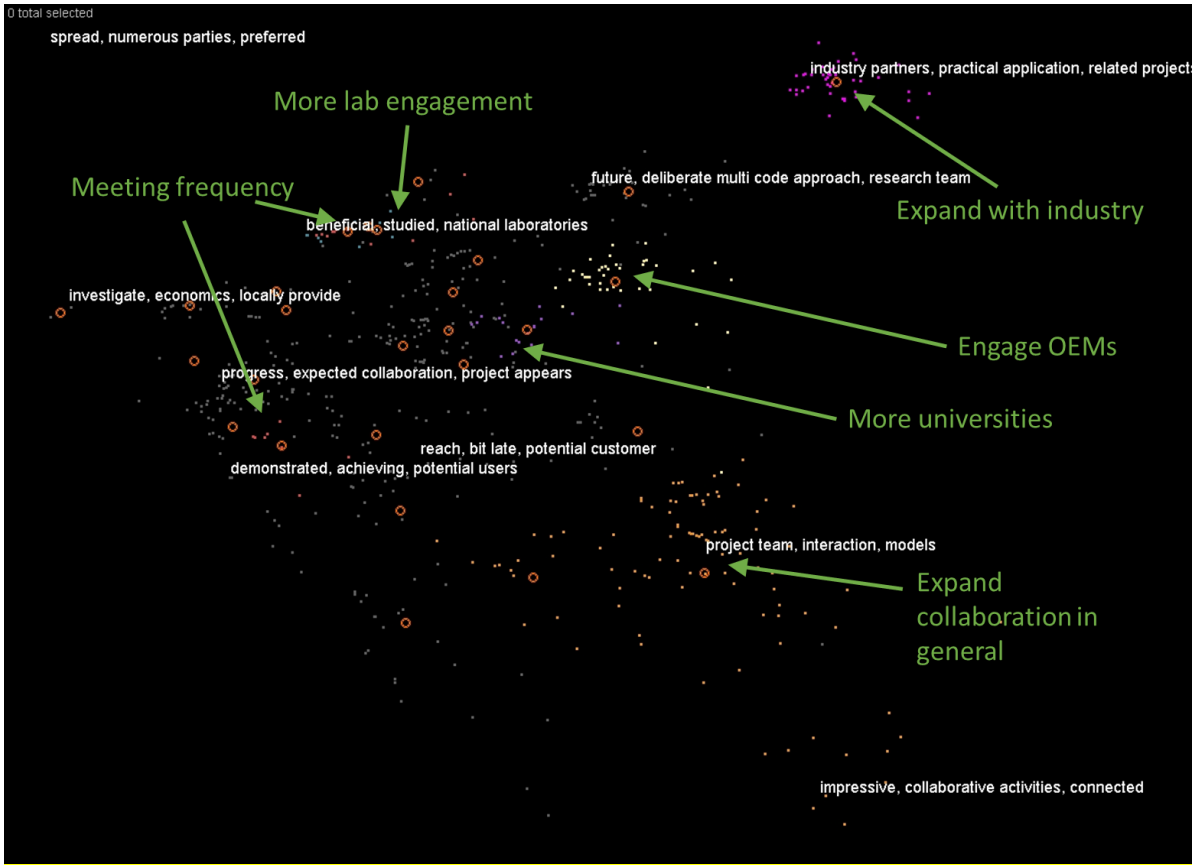

Figure 14. This galaxy visualization from IN-SPIRE™ shows the different clusters of comments containing recommendations related to a project's collaboration and coordination.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 745 records that contained a recommendation for project's collaboration and coordination:

18% of the records were related to general recommendations for expansion of collaboration:

- “A broader, more substantial list of active collaborations would be expected”
- “Project would have benefited from incorporation of an industrial partner”
- “Suggest expanding the list of hose providers”
- “Additional collaboration with utilities is recommended”

8.3% of the records were related to more engagement with OEMs:

- “Project would benefit from increased interaction with other industrial partners and OEMs”
- “This is one area where they could benefit from more industrial interacts”
- “The project would benefit from having a [ ] manufacturer either as a partner or consulting role”
- “The project would benefit by engagements with equipment manufacturers and OEMs”
- “This reviewer thinks the project needs more collaborators. In particular, the team should collaborate with experts in modeling, vessels, heat exchangers and material thermodynamics”

5.8% of the records were related to expansion of industrial partners:

- “This reviewer would recommend identify the top five leaders in the (specific industry) to solicit opinions of the work performed to date and recommendations for ongoing work”
- “Collaboration with an industry partner is suggested”
- “It would be advised that collaboration with [ ] is conducted to expedite the progress of this project”
- “I would like to see an industry partner here to ensure that the technologies will have a commercialization path”
- “No industry partner is stated. It might be a good idea to have industry inputs during this activity”
- “As the project started recently, it is suggested to include an industry partner to help in productionizing this technology”

3.5% of the records were related to recommendations associated with meetings and meeting frequency:

- “Monthly meetings are barely enough to keep this complex project running smoothly… I encourage weekly web-based meetings and monthly face-to-face meetings in the next year”
- “I suggest showing the frequency of web-based meetings and face-to-face meetings to clearly describe the collaborations”
- “I would like to see a matrix of when and who meets by web-based meetings and face-to-face meetings to better understand the collaborations”
- “It is recommended that regular meetings are held to improve the effectiveness of the integration activities”
- “With such a large team, it would be useful to remind the audience how the work was coordinated (meeting schedules, internal reviews, etc.)”

2% of the records were related to more collaboration with universities:

- “It would be better if rapport could be established between [ ] and the University of [ ]”
- “There seem to be some missed opportunities. For one, I would think collaborations with academia would be possible”
- “But I suggest more collaborations with US universities. All the academic collaborations that the PI presented are with universities outside US”
- “A pretty broad team has been assembled including DOE labs, several engine makers, and fleet operators - add a couple of universities and it would be perfect”
- “I suggest also interacting with other universities that have common research interests”

1.3% of the records were related to more collaboration with national labs:

- “I might suggest adding an expert in (technology name) as there is significant expertise at the national Laboratory level”
- “I would suggest collaborating with a DOE lab that develops software”
- “I would suggest trying to engage with a national laboratory team”
- “It would be better if rapport could be established between [ ] National Laboratories and [ ]”
- “The future collaboration with [ ] is good, but I strongly suggest collaboration with the national labs to work on cost modeling”

6.4 Detailed Results of Relevance Recommendation Themes

This section contains the detailed results of IN-SPIRE's analysis of recommendation themes contained in Relevance comments. This dataset contained 445 individual records. Statistically, IN-SPIRE™ organized these comments into 17 clusters, which distilled to 7 interpretable themes. Figure 15 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

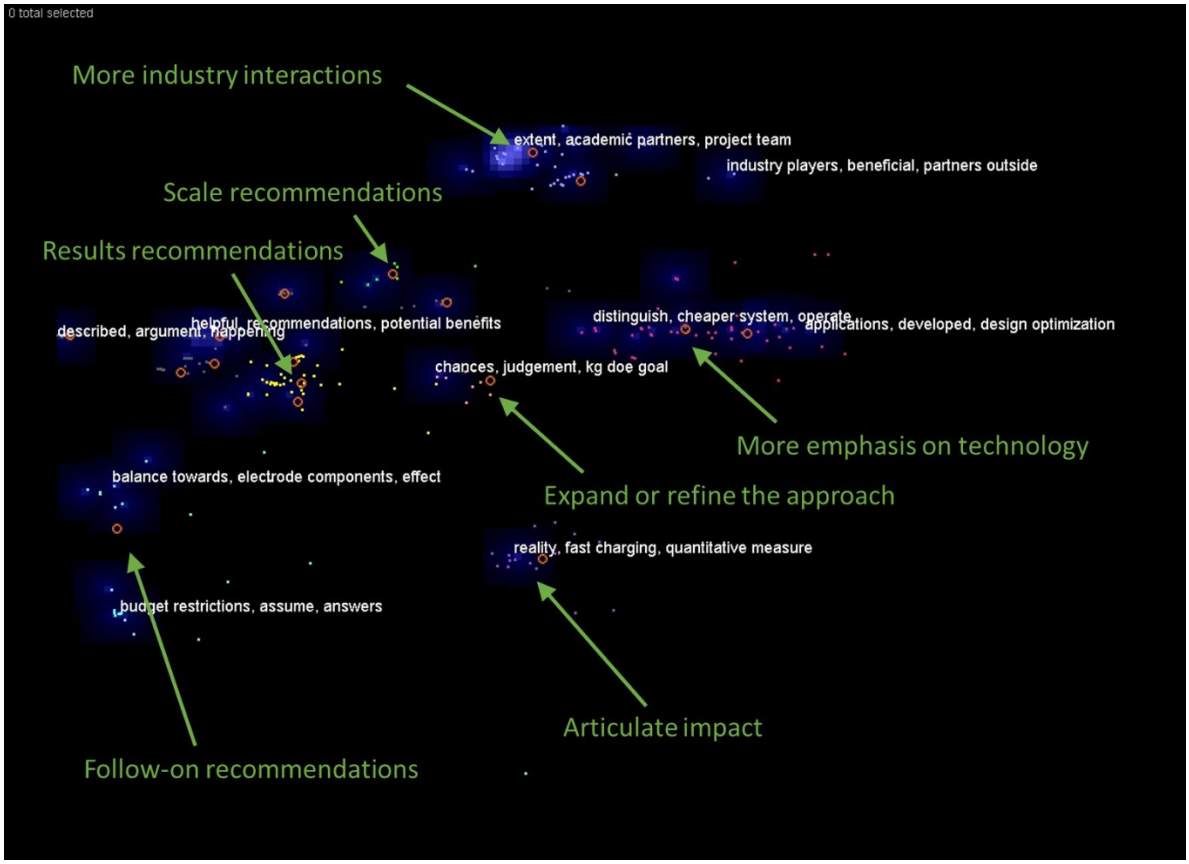

Figure 15. This galaxy visualization from IN-SPIRE™ shows the different clusters of comments containing recommendations related to a project's relevance.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 445 records that contained a recommendation for project's relevance:

16.2% of the records were related to general follow-on recommendations*:

- “I suggest the creation of a library of models (like a database of models), available to others for direct application”
- “It would be fantastic if the outcome of this project could support commercial development on many (any) system”
- “Recommend the project consult with [ ] on this aspect and get their input”
- “It would be ideal if this equipment could be purchased or transferred to [ ] so that additional work in this area could be completed by the national laboratories”
- “This project needs quantitative milestones in order to be of more relevance to the DOE”

12.4% of the records were related to recommendations for more emphasis/deeper information about the technology:

- “It would be good to highlight the importance and benefit of this particular FC technology”
- “It would be good to establish that first, and to determine if this technology will afford a winning solution for the future as compared to current OEMs technology paths”
- “It would be useful to see more discussion on how this would be accomplished, and what limitations or boundaries this would impose on deployment of this technology”
- “I would personally like to see a company or end user that would like to develop this technology to buy down risk on the application side”
- “I think a comparison with advantages of this approach to a reforming methanol fuel cell would be valuable. As would a suggested use case of the technology”

9% of the records were related to recommendations pertaining to availability or description of results information:

- “But this project relies too much on results of this specific round of tests. It would be better to include continued investigations on real world stations”
- “I recommend including graphs or tables that can get at trends over time, perhaps with a sliding 3-year average of results”
- “It would be good to see how those studies are being done and the results”
- “It would be helpful to understand how the results will be used”
- “To understand the degree of progress, it would be useful to include information about comparable results from past years or the start of the project”

8.5% of the records were related to recommendations to expand industry interactions*:

- “As feedback, I can only suggest the team consider periodic checks on the relevance of the consortium's tasks to industry”
- “It would be encouraging to see a major industrial end-user on the team”

* While this is a recommendation theme found in relevance comments, it seems less associated with the relevance criterion and more with a different criterion. Future work could align these themes with a more appropriate criterion

- “It would be good for the team to reach out to industry partners to get feedback on next steps and strategies for strain development”
- “It will be useful if the collaboration is extended to the industry, so the team can benefit from the industrial viewpoints”
- “I recommend evaluating how the main project deliverable, the CDP, is being publicized so that people in the industry can find it”

4% of the records were related to recommendations to better articulate the impact of work performed:

- “It would be good to attempt to quantity the impact these various solutions can have on the immediate defect being investigated”
- “It would be preferable if there were some approximate idea of the impact it could have now, to determine whether it was in theory worth pursuing”
- “The relevance and impact would be clear if the new technology is compared against other reported PI alternatives”
- “Would be good to add a calculation relating the impact of meeting the goals with the effect of petroleum displacement”
- “It would be applicable across all vehicle types, so potentially could have a big impact”

2.2% of the records were related to expanding or refining the approach[*]:

- “It would be interesting to know if there are further applications for this approach”
- “The goals and impacts of each node are stated but they could benefit from a SMART approach”
- “Even if this was a feasible, well performing program (which it currently is not), I would question this approach”
- “I would expect OEMs to take a 'optimize for my vehicle first' approach”

1.6% of the records were related to recommendations about scaling:

- “Also, it would be helpful if the choice of co-products were vetted by industry and economists to ensure relevance for scale-up”
- “I would prefer that the final aspiration in terms of scale of the final version of this reactor were made clearer”
- “It would be useful if the efforts can be used to scale-up the temperature of operation”
- “I would be keenly interested in understanding the potential for these cells to scale up to a 120+ kg/h at 70 MPa output pressure which would be relevant for where the industry is going with high demand regions”
- “I think it would be valuable to get a sense for how this will be deployed in the real world and at scale”

[*] While this is a recommendation theme found in relevance comments, it seems less associated with the relevance criterion and more with a different criterion. Future work could align these themes with a more appropriate criterion

6.5 Detailed Results of Future Research Recommendation Themes

This section contains the detailed results of IN-SPIRE's analysis of recommendation themes contained in Future Research comments. This dataset contained 1,328 individual records. Statistically, IN-SPIRE™ organized these comments into 36 clusters, which distilled to 7 interpretable themes. Figure 16 shows the IN-SPIRE Galaxy visualization for this dataset, showing each cluster of records as a different color. Each dot represents a record extracted form a reviewer comment, and the orange circles represent the cluster centroids. The white labels are IN-SPIRE's labels to differentiate different topic areas in the visualization, and the green labels were added by PNNL to describe the major theme contained within each cluster of comments.

Figure 16. This galaxy visualization from IN-SPIRE™ shows the different clusters of comments containing recommendations related to a project's future research plans.

The following are examples of specific reviewer comments taken from the records. Percentages shown below are relative to the 1,328 records that contained a recommendation for project's future research plans:

17.7% of the records were related to recommendations for further validation or refinement of the results for additional use cases:

- I suggest the project test how well their outdoor results in a small outdoor pond on a controlled site translates to larger ponds that are operated as would be for a large commercial facility
- Therefore it is recommended that future work includes a comparison of the results (cost and weight) with [ ] alloy
- Nevertheless, I would propose that the results of the [ ] should be summarized in a book as well
- There is nothing in the presentation suggesting that results will be available in any other form, or disseminated more widely
- But it would be good to get the same results for the [ ]

11.2% of the records were related to recommendations encouraging working with others avoid duplication of work:

- It is suggested though in the future to disclose more of progress toward goals and how the project will enable better bio-based product once they feel more comfortable to disclose this publicly
- I recommend a schedule reassessment be performed by 4Q17 and published
- It would be great if [ ] National Laboratory also published reports that compare all the different charging systems in one report
- Electrolyte formulation recommendations for specific cathodes, i.e. [ ] and [ ] have to be published
- I would suggest that this workshop and the other two educational projects have close collaboration in terms of contents, examples, etc., to avoid duplications

6% of the records were related to recommendations for expanding the plan for use of previously collected data:

- It would be useful if consistency tests for collected data could be developed
- It would be good to hear more about how the data will be used to control the process though, and how advanced control systems will be developed
- Usability of the tool is second only to the quality and quantity of data to ensure the success of the data
- It would be recommended to include work with system/unit cell modelling efforts to provide either validation data or directed testing to further improve the model mechanisms or predictions
- Would be good to see more [ ] test data by next AMR
- It is recommended that they compare their data with some more traditional baseline alkaline electrolyzer catalyst data

4.9% of the records were related to recommending the team to focus their efforts in other directions or further engaging with industry:

- I recommend increased collaborative interactions with the [ ] and [ ] consortia to provide broader and more detailed characterizations

- This part of the team may need some reinforcements
- Again the team is suggested to get more outside partners involved, particularly if they can bring analysis capability onboard
- I would encourage the team to engage with [ ] R&D teams
- The team might benefit from discussions with a range of industrial experts to understand better hydraulics challenges, define goals and metrics before moving to future work

3.8% of the records were related to recommendations for more focus:

- I would suggest a re-focus (or re-definition) of the project
- Would like to see more justification for choice of particular directions, given the huge set of possible focus areas within this area
- Prioritizing of the work is needed. I don't believe that all goals can be reached in this timeline. I would try to focus more
- I would recommend that they focus less on DFT and powder scale up and more on electrochemical methods (RDE and CCMs)
- It would be preferrable if the project focusses on these materials rather than…
- Those that include precious metals in particle cores

2.1% of the records were related to recommendations for scaling:

- The reviewer would like to see more evidence or thoughts regarding scale-up
- I suggest incorporating more reliable/reproducible characterization of screened phenotypes (lipid accumulation, improved growth, salt/temperature tolerance) to boost confidence that there is an actual improvement prior to deploying at scale
- Carry out the engineering scale-up
- Scale-up of the catalyst is needed, but it would be helpful to have a supplier involved. Perhaps this could be addressed in a later funding period
- It may be early in the program, but would like to see additional discussion on the cost of a full-scale system

2% of the records were related to recommendations to make adjustments to phases:

- Would like to see a module built and tested in next phase
- The second phase survey seems warranted, and it would be good to try to tease out company’s actual planned hiring from aspirational goals
- I would put MEA testing off until the Phase II
- Given the delays in Phase I, I suggest assuming a successful build that the length of the demonstration of Phase I hardware be shortened and the number of vehicles in Phase II be reduced by a factor of two
- Analytic models of the controls based on state variable and similar mathematical methods in the next phase is expected

7.0 Discussion

Overall, we found IN-SPIRE to be a useful tool in identifying themes within these reviewer comments. There is a high value in understanding the themes that exist in these kinds of reviewer feedback and comments, where it is not feasible for someone to read each and every comment. IN-SPIRE was an effective tool in helping to navigate around the large corpus where we could analyze content at a macro scale and yet drill down into subsets to analyze content at a more focused scale.

Through the analysis of over 42,000 reviewer comments across multiple EERE R&D programs, this IN-SPIRE analysis enables EERE R&D managers to learn lessons to aid project management based on information that extends beyond their own experiences.

7.1 Value of the Analysis to R&D Managers

We hope that the themes identified will be helpful to EERE R&D managers and can help inform their decisions about their future projects or future peer-reviews. Knowing the strengths and weaknesses of previous projects can provide the EERE R&D managers the information to help guide their management of future projects towards success and steer them away from failure. The lessons learned from this study will help R&D managers proactively deliver recommendations to these projects throughout their lifecycle, to apply and learn lessons from previously executed projects.

The strength themes that occurred most frequently in reviewer comments indicated that it was important to reviewers that projects have engaged collaborators balanced among national labs, industry and academia, with well-defined roles for each collaborator. Projects should have benefits to industry and a strong potential for technology transfer. Strong projects include tasks to develop new models, improve previous modeling activities, or contribute to new material development. Improved cost efficiency was important, and projects should highlight their plans to improve cost for technologies or industrial processes. Goals should be clearly stated, along with progress towards achieving these goals, in both current and future work plans. Challenges and barriers should be identified, with plans aligned towards overcoming them. R&D managers can promote projects that have these same qualities.

The weakness themes that occurred most frequently in reviewer comments indicated it was concerning when there was a poor connection to industry and perceived difficulties for technology transfer, including challenges transitioning from research to industry scales. Reviewers were also concerned when there were challenges with the team, poor interactions among the team members, and missing collaborators. Reviewers were also concerned with the misalignment of projects to program goals, projects that were failing to meet their own targets and goals, and projects with future plans that were too broad, lacked focus, or didn't adequately address industry challenges. R&D managers can steer their portfolio away from these kinds of projects or provide project oversight to preempt the occurrence of these potential weaknesses.

The recommendation themes that occurred most frequently in reviewer comments indicated that reviewers would like to see more performance data, more cost assessments, more comparative analysis with other approaches, and more comparisons with the current state of the art. Reviewers recommended expanding collaborations to include a balanced team of industry, academia, national labs, and consortia members. And reviewers wanted to see more details on

the impact the project is having, and were also interested in understanding how technology could have additional applications in other domains or subdomains.

7.2 Value of the Analysis for Further Improving the EERE Peer Review Process

The themes identified as part of this analysis can also be helpful for further improving the EERE peer review process. While many of the identified themes related to the projects themselves, a number of themes were directed at the presentations. This IN-SPIRE analysis identified five opportunities for EERE to further improve its peer review process to enable reviewer comments to be more readily interpreted and used within the IN-SPIRE analysis or without the analysis.

1) Obtaining quality review feedback is essential for these projects to maximize their impact. One of the themes observed as both a weakness and a recommendation was related to challenges faced by reviewers in being able to provide their feedback. Reviewers struggled with understanding and evaluating portions of the projects because of two main areas: information missing from the presentations, and provided information being unclear and difficult to comprehend. Reviewers should be encouraged to present comments about missing or unclear presentation information separately from technical comments related to the actual project specific material.

Peer Review Process Improvement Opportunity:

- A potential improvement to the peer review process would be for EERE to add an additional question in its guidance to elicit reviewer feedback directly related to the quality of the presentation materials, so quality-related kinds of weakness and recommendation comments can be responded to separately from core themes pertaining to the 5 primary review criteria.

2) As described earlier in section 3.0, reviewers often make observations that lack a judgement or are too subtle to interpret their judgement.

Peer Review Process Improvement Opportunity:

- Offices could add guidance to peer reviewers to encourage reviewers to include a judgement whenever they make an [important] observation. The guidance could provide reviewers with several examples when an observation statement is too subtle to integrate their judgement.

3) Every reviewer expresses themselves differently and uses language differently. As such, some reviewers are not always clear in expressing strength, weaknesses, and recommendations. Also, sometimes reviewers mix commentary for multiple review criteria together.

Peer Review Process Improvement Opportunity:

- A recommendation is to explicitly ask the reviewers to identify strength, weaknesses, and recommendations for each criteria, and to provide distinct space for reviewers to

provide this information. This will take the guesswork out of determining if a given comment should be interpreted as a strength or a weakness or a recommendation. For example, if reviewers are provided with a template such as the following, they could be asked to organize their review response according. This does not mean reviewers must fill out a table per se, but they are to be encouraged to clearly indicate when a response is a strength, weakness, or recommendation. This guidance would also help address the problem of some reviewers mixing commentary for multiple criteria together. It would help could provide for more accurate mappings of comments to criteria.

Name of project: _____ Date: _____			
Criteria	Strength	Weakness	Recommendations
Accomplishments and progress			
Approach			
Collaboration and coordination			
Relevance			
Proposed Future			

4) Reviewers do express their positive and negative opinions. However, they don't always explain their reasoning behind the positives and negatives. For example, a comment like "This project has major problems. It should be drastically restructured or, preferably, canceled" indicates a negative viewpoint from this reviewer without including any specific weakness attributes that led to this viewpoint.

Peer Review Process Improvement Opportunity:

- A recommendation is to include explicit guidance to the reviewers to more carefully identify the strengths and weaknesses in addition to positive and negative attributes of a project.

5) Some reviewers write in a style that is very confusing. Some reviewers number each theme within their comment, while some separated them with hyphens, semicolons, or periods. For example, rather than a reviewer writing "This project has engaged collaborators in industry and is working well with OEM manufacturers to reduce their production costs for materials that will exceed program targets", the reviewer should split this sentence into three sections: one highlighting a strength in collaboration, one highlighting the impact on OEM manufacturers, and one highlighting exceeding program goals. These sections could be individual sentences, or they could be separate paragraphs within the same comment block. Specific guidance should specify how to separate these sections, such as using blank space between paragraphs.

Peer Review Process Improvement Opportunity:

- A suggestion would be to supply guidance to reviews on how to encode multiple themes within a single comment, to ease the burden of parsing them.

7.3 Challenges

We faced six challenges during our analysis of the reviewer comments. We discuss each one below, how we handled the situation for our analysis, and offer suggestions for potential solutions to these challenges in the future.

Challenge 1: Validating why Strengths outnumber Weaknesses and Recommendations

After initially observing the larger proportion of strength comments to weakness and recommendation comments, we took a closer look at the data and our process to validate the accuracy of this finding. One obvious reason is that reviewers simply included more strengths than weaknesses or recommendations in the review comments. In addition to this explanation, we came up with additional explanations that could have been the result of the process we used in the IN-SPIRE analysis.

One reason the strength comment counts were so high is because we included some very generic words, such as “good”. A word like “good” could be used in many different contexts, and not always in the context of a strength theme. Of the 24,972 records containing a strength keyword, 9,808 of them contained the keyword good.

Another reason the strength comment counts were so high is because it can be challenging to detect negative concepts, like weakness, because in English, we often express negative concepts by negating a positive word. Some of the occurrences of “strength” keywords could actually be “weakness” concepts based on the context in which they are used. For example, the phrase “the project is failing because collaboration is not good” is more likely to occur than “the project is succeeding because collaboration is not bad”. If we encountered a statement like “the project is failing because collaboration is not good”, this would have been counted in our strength counts, but ultimately not associated with a strength theme – see Challenge 5 for more information.

We acknowledge that our lexicons were very broad, perhaps overly broad. Our goal was not to fine tune a lexicon to find strengths and only strengths, but to lead us towards a majority of records containing strengths that would cluster together based on other keywords. We intentionally cast a wide net, then use the capabilities of IN-SPIRE to locate those clusters that could be interpreted into themes. Future iterations of this kind of analysis may want to consider further refinement of any lexicons to match their specific analytic tools and workflows. Some examples of some keywords we might consider removing in the future include: remove “good” from the strengths lexicon, and remove “challenge” from the weakness lexicon.

Challenge 2: How to handle reviewer comments with no lexicon terms

As stated earlier in this report, we observed a substantial number of reviewer comments did not contain any keywords from our strengths lexicon, weaknesses lexicon or recommendations lexicon. We studied this collection of reviewer comments to make sure we were not excluding any major themes. We found this collection largely contained statements where a reviewer was stating a fact, or a statement where a reviewer was offering an observation without making judgement.

Every reviewer expresses themselves differently and uses language differently. Our process for lexicon development included many iterations of trying to include as many terms as we could find in the data. Our process was not as scientific as it could have been for determining the limit, but our instincts informed us when we were reaching diminishing returns by continuing to iterate. This is an area for improvement in future analysis activities.

Challenge 3: Reviewer comments in one criteria that appear to address aspects of another criteria

We saw a lot of cross-cutting themes, and some themes that seemed to not belong in the criterion for which the comment appeared. We also found comments like “please see above” and “many, as discussed above”, that indicates some reviewers entered their comments in a single text field but addressed multiple criteria at once, then referenced their previous comment in the field for a different criterion. We did not attempt to re-align these themes with an alternate criterion, nor did we attempt to cross-reference comments that referred to other comments. We relied solely on the information provided to us in the spreadsheet of reviewer comments to align comments with criteria.

Challenge 4: Reviewer comments that contain a strength and/or a weakness and/or a recommendation in the same comment

As stated elsewhere in this report, we found a fair number of comments that contained strength keywords, weakness keywords, and recommendation keywords inter-mixed in the same reviewer comment. In each case where this occurred, the same comment was analyzed for each concept the comment seemed associated with, so a comment that contained a strength keyword and a weakness keyword was included for strength theme analysis and for weakness theme analysis.

Challenge 5: Lack of a “theme” to accompany a lexicon keyword

We found the reviewers are not shy about emphasizing the positives and negatives of the projects they are reviewing. We discovered this while performing sentiment analysis on the dataset. But we found that reviewers don’t always explain their reasoning behind the positives and negatives. Since our goal was to identify the themes of strengths and weaknesses, this additional context is helpful in identifying the strength or weakness that was the cause behind the positivity or negativity. For example, with these comments, the reviewer thinks highly of a project, but doesn’t express the rationale or “strength” of the project that contributes:

- “I felt that some of the delays were unnecessary and could have been managed better”
- “This has great relevance within the composites community. These are problems that need to be attacked.”
- “The performance achieved is low and inconsistent”
- “The work is unlikely to yield large reductions in pipeline costs”
- “In my opinion, this is the best project yet presented.”

The same can also be said of weakness comments – sometimes a review expresses a negative attribute for a project but fails to express the weakness this is tied to. But with small additions of a few more words, these positive and negatives can be correlated with strength and weakness themes.

Challenge 6: Confusing writing styles

Some reviewers write in a style that is very confusing for our lexicon-based approach for identifying strengths and weaknesses. None of these comments are weaknesses, some are actually strengths, but they only include words from our weakness lexicon:

- “No indication that work planned was not completed due to lack of resources.”
- “It would be difficult to identify weak points of this project.”
- “I do not have any comments on project weaknesses at this time.”
- “Could not find any weaknesses”

When encoding multiple themes in a single comment, there was a variety of writing styles to separate these themes within a single comment. Some reviewers numbered each theme within their comment, while some separated them with hyphens, semicolons, or periods. Although we separated on sentence boundaries, we ended up using a variety of characters, symbols and patterns as delimiters, including embedded newline characters, periods, question marks, semi colons, a number followed by a period followed by whitespace.

7.4 Future Work

The process of performing analysis on the reviewer comments has identified areas where the data collection and preparation and IN-SPIRE analysis can be further improved.

An area of future work would be to expand this analysis to include additional data. 20,662 comments in the provided database were excluded from our analysis because they did not map to one of the 5 designated criteria; future work could seek to place these comments into one of the 5 criteria so they could be included in the analysis results. Methods developed to assign these comments to a designated criteria could also be applied to mis-assigned comments to better align them with a more appropriate criterion.

Another area of future work could use some of the additional data fields that were not populated in our database, such as project funding levels, project duration, or project stage at time of review, such as new, ongoing or sunsetting. Comparative analyses can be performed to look at the differences in themes identified for different categories of projects. For example, a theme we saw in the weaknesses of Accomplishments was generally labeled as “difficult to evaluate” with some specific excerpts including “difficult to judge, as the project is only 3 months old” and “project has just begun so hard to judge the accomplishments”. Perhaps projects that are new should be reviewed and evaluated differently from projects that are more mature. If different themes emerge for different categories, this could inform EERE R&D managers to adjust their guidance for managing different types of projects.

A follow-on activity to this analysis activity would be to apply “ground truth” data for knowing if strength themes are actually associated with successful projects, or if weakness themes correlate with failing/struggling projects. Our analysis was based solely on the presence of strength, weakness, and recommendation keywords, without any concern for the health or status of the actual project the comments were associated with. A future analysis model might incorporate this criteria, so only strength themes of successful projects are highlighted as positive qualities, or weakness themes of failing projects are highlighted as negative qualities.

A future analysis activity could also look beyond the statistically based algorithms that IN-SPIRE relies upon and consider the use of Artificial Intelligence/Machine Learning (AI/ML) algorithms

for performing Natural Language Processing (NLP) to recognize strengths, weaknesses, and recommendations. AI/ML models can be trained on labeled data, where someone has curated a set of reviewer comments to indicate the presence of a strength, weakness, or recommendation, and the AI/ML model can then predict the presence of these qualities on a new corpus of reviewer comments

Reviewers play a key role in determining the success of EERE projects by offering their expertise. EERE R&D managers can benefit from these reviewer comments when managing their portfolio of projects and guiding their projects towards a successful completion. Recognizing the strengths, weaknesses, and recommendations offered by reviewers will aid the R&D manager in providing adequate guidance to the project team, and in making subsequent funding decisions.

Appendix A – Method

Dataset Preparation

When PNNL first received data, we performed some data exploration tasks to get more familiar with the data. Since the goal was to extract meaningful themes from comments, we wanted to make sure the comments contained enough content for IN-SPIRE™ to be effective. We performed some basic analysis to look at the distribution of comment length in terms of character count, word count, and sentence count. We found the following:

- There were 2208 unique projects
- Five projects had over 200 comments associated with them
- Twelve projects had only two comments associated with them
- The longest comments had almost 10,000 characters
- Three comments had over 1000 words in them
- 210 records had comments containing only a single word
- Over 3000 comments contained five words or less
- Roughly 20 comments contained over 30 sentences.

Based on these numbers, PNNL hypothesized that the short comments, which included “ok”, “no”, “unknown”, “cut” and “outstanding”, would likely not provide a lot of value because there was simply not enough content to determine the how or why behind the comment. And PNNL hypothesized that really long comments could prove to be challenging to cluster with other comments because they likely contain multiple thoughts within a single comment field.

After discussions with the sponsor, we subsetted the original dataset to only include comments from R&D comments. Our hypothesis was that comments for non-R&D projects would add more noise to the dataset and make the comments for R&D projects harder to locate. PNNL performed a quick analysis using one of the criterion as an example, and we found the clusters and themes found in IN-SPIRE™ were fairly consistent between the larger and slightly smaller datasets.

While getting acquainted with the data, we noticed that some comments were very rich in content, such that an individual comment contained multiple concatenated thoughts highlighting strength, weakness, and recommendation in the same comment. We evaluated how well IN-SPIRE™ might perform if we split these apart, and we found that IN-SPIRE™ performed well when we split comments based on sentence boundary. Because many comments contained multiple mini-comments within one block of text, we found that by splitting these apart, common mini-comments tended to organize into more coherent themes.

For the remainder of the method discussion, we use the following terminology:

- Comment: an individual row from the sponsor as provided in the spreadsheet
- Individual thought: An individual strength, weakness, or recommendation contained within a comment
- Record: An individual sentence extracted from a comment

- Document: The IN-SPIRE™ unit of analysis, which were individual records for our use case
- Concept: Strength, weakness or recommendation
- Keyword: A single word or multi-word phrase identified by IN-SPIRE™
- Theme: A human interpreted meaning of an IN-SPIRE™ generated cluster of records
- Lexicon Axis: A set of keywords we curated to describe an individual concept.

PNNL created a second spreadsheet that was derived from the original spreadsheet where we split each sentence of a comment into its own record. This resulted in a spreadsheet of 110,309 records. Splitting on sentence boundaries was not precise but we were willing to accept certain margins of error. For example, we split on the period character followed by a blank space character, but some comments contained a period in the middle of them when abbreviations were used. Some comments contained odd combinations of whitespace or other characters in between sentences so they did not get split properly. Some reviewers also inserted bullets or separated individual thoughts with other notations – many of these were omitted by our splitting mechanism - these comments remained intact with their original notation. Individual sentences extracted from a comment were added to a new PartialComments field and the original comment remained in the Comments field. A full data dictionary of our modified spreadsheet has been provided in Appendix B.

We then used IN-SPIRE™ to create a single dataset of all 110,309 records. We did not use this dataset directly for analysis purposes, but this was our starting point for all subsequent analyses. We created our dataset using the PartialComments field as our only Computational field. The Comments field was present added as a Regular Text field but was not marked Computational. This allowed us to see the context of how individual sentences appeared in a larger comment without influencing the IN-SPIRE™ algorithms. Other categorical data types were added as Regular text fields and Categorical but not Computational, to enable quick filtering via the Facets tool. The PartialComments field was also denoted as the Title field to make navigating documents via the Document Viewer easier. We did not select a Date field for our dataset. Appendix C contains more information about how we constructed our IN-SPIRE™ dataset

As we approached analysis of records for each criterion, we used the Facets tool to select all of the records for a given criterion, then used the “Subset from Selection” function to create a subset dataset. Subset datasets contain some but not all of the records from the original dataset, and enabled IN-SPIRE™ to utilize previously computed artifacts, such as full-text indexes, more efficiently. Subsets also allowed for custom stopwords and other dataset parameters to be specified, providing a capability for different criterion to use different stopwords lists. IN-SPIRE™ performs better as an application with datasets that are smaller in size.

Lexicon Approach to Subsetting by Strength, Weakness, Recommendations

PNNL used the Lexical Analysis tools within IN-SPIRE™ as the primary tool to identify records of interest. Lexical Analysis is typically used to score documents along different axes, where each axis is a weighted list of terms associated with a given concept. Lexical analysis is often used with sentiment analysis to measure the positivity or negativity of different blocks of text. Since the goal of our work was to identify themes associated with strengths, weaknesses and

recommendations included in records, we constructed a lexicon that encompassed these concepts, with each concept as its own axis. We then measured the subset of records for a given criterion against the lexicon to find records that were associated or not associated with the axis concept. We were less concerned with the lexical scores of each record, but more concerned with the presence or absence of a lexicon concept for a given record. This was a key step in our process for identifying records that were associated with a given concept.
Our general approach consisted of these major steps:

1. Create a subset dataset for all records associated with a criterion, measure Lexical Analysis with our custom lexicon
2. Perform Lexicon Expansion to find additional keywords to add to our lexicon concepts
    a. Segregate all of the records that didn't score against any concept into an IN-SPIRE™ group
    b. Examine the summary terms and cluster labels for these documents and look for candidate lexicon terms to add to one of the axes
    c. Re-run lexical analysis if necessary
3. Perform concept-specific analysis to identify themes
    a. For each of the concepts, use IN-SPIRE™ to focus on the records that scored in that concept
    b. Explore clusters and keywords identified by IN-SPIRE™ to discover themes associated with this concept

A few of these steps are described in more detail below.

Lexicon Generation and Expansion

Initially, we performed internet searches to look for pre-existing lexicons for strengths, weaknesses and recommendations, but we did not find anything that looked suitable for our purposes. We ended up creating the term lists for our lexicons manually, by assembling words that captured the essence of the concept. We used two techniques for finding additional candidate terms to add to our lexicon. One technique used an elimination technique to focus on or eliminate the known words to see what words we were left with or what words were highly related to our focused words. This technique worked well for immature lexicons. The other technique was to look at the top terms in the documents that were excluded from our current lexicon lists, which worked well for mature lexicons

In the first technique, we would take our seed lexicon terms, search for them in our dataset, then use the IN-SPIRE™ Focus and Ignore feature to focus and look for related terms or ignore and look for other dominant terms. We would un-ignore certain result sets and focus on others to see what words were related to the different sets of words we were searching for. While this process seemed a bit haphazard, it resulted in some fairly good term lists. We would often include the word of the criterion itself in our searches to tease out different adjectives that reviewers had used when describing the criterion.

In the second technique, which was performed when our lexicons were more mature, all of the documents that did not contain any term from the lexicon were placed into an IN-SPIRE™ group, and the Focus feature of IN-SPIRE™ was used to focus on these documents. By selecting all of these documents, the Summary tool shows the top keywords for these documents. We would use the IN-SPIRE™ Focus feature to drill into a single keyword at a time and inspect the sub-clusters that form. We would also spot check documents in different clusters to see if other words were used that were outside of our lexicon. This was a manual

process, but one that went fairly quickly. During this process, we discovered additional lexicon terms for the “relevance” criterion.

Once we were satisfied our lexicons were appropriate for the criterion being measured, we performed one final lexical analysis measurement on the comments for a given criterion. We used the Scores tab to select all non-zero documents, then created a subset dataset for each lexicon axis, resulting in three additional datasets per criterion: a strengths dataset, a weaknesses dataset, and a recommendations dataset.

Theme identification

The process above described how we created and used lexicons to find those documents containing strength, weakness, and recommendation concepts. The next step in the process was to look at all the documents that scored non-zero values for a given axis and look for commonalities among the comments to identify themes. IN-SPIRE™ provides capabilities to streamline this process. We relied on the Galaxy cluster labels, Galaxy cluster centroids, terms in the Summary pane and the Charged Terms tools. Here are some of the key steps we used:

- After creating a subset dataset for a given criterion, measuring with the lexicon, and ignoring all of the zero-score documents, we used a series of techniques to identify terms that could potentially be a part of key themes of the dataset:
- Next, we Open the Galaxy tool (Figure 17) and examine the terms that IN-SPIRE™ selected for cluster labels. Although these labels are often more discriminatory than descriptive, they provide hints to how different areas of the Galaxy may differ from others. The probe tool (shown with the green arrow in Figure 17) can show labels elsewhere in the Galaxy visualization tool. The orange circles in Figure 1 show the cluster centroids for each IN-SPIRE™ cluster and can be used to select all documents within a single cluster.

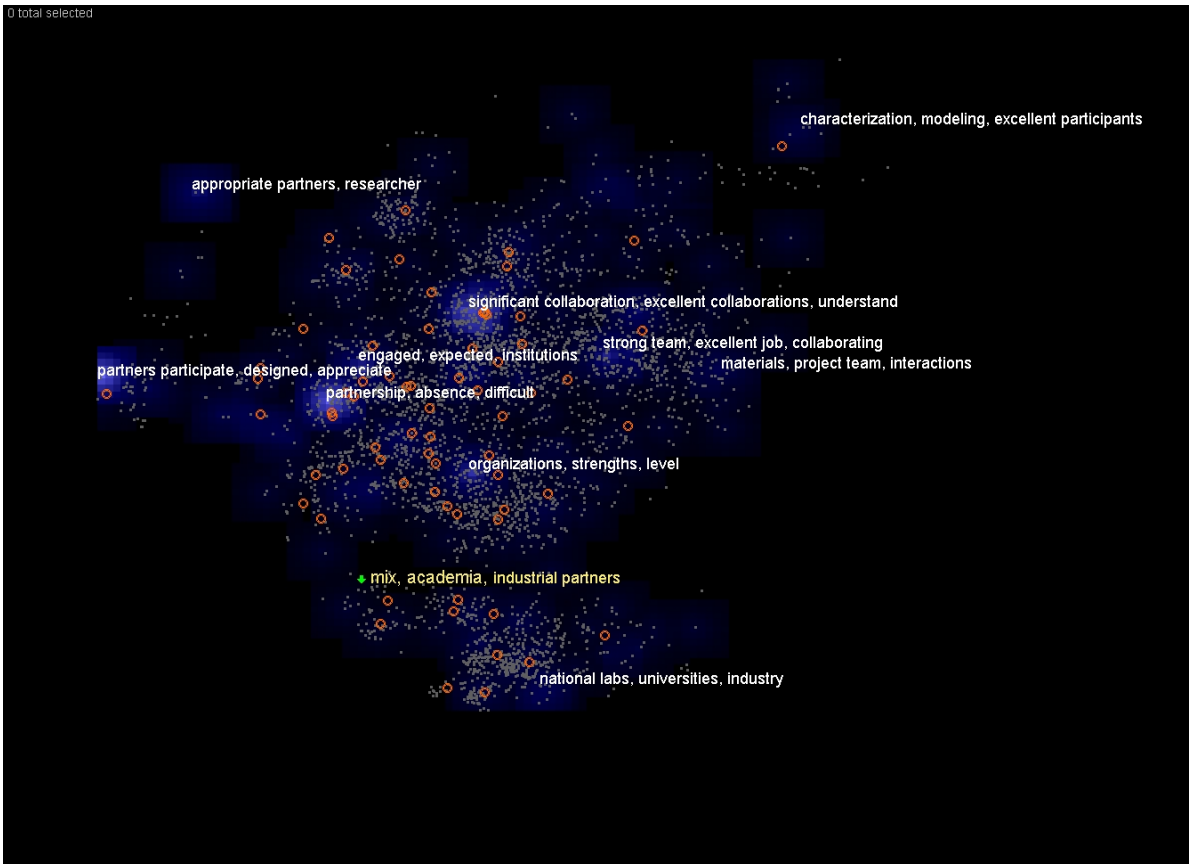

Figure 17. Labels in the Galaxy View with Cluster Centroids and the Probe Tool.

- Using the Galaxy tool, select either all the documents or select a cluster of documents, and use the Summary Tool to see the top keywords contained in the selected documents (Figure 18). Also use this list to identify candidate outlier keywords.

SUMMARY

Term	Contribution	Docs
beneficial		32
project team		23
interactions		10
expertise		6
strong team		8
analysis		4
addition		4
assembled		3
modeling		2
users		3
researchers		2
contribution		3
progress		2
teams		3
job		2
excellent collabor...		2
companies		2
effort		2
labs		1
engagement		2
involvement		3

Figure 18. Summary View within IN-SPIRE™

- Using the Galaxy tool, utilize the cluster centroids to select each of the cluster centroids, and use the Summary Tool to see top keywords in a cluster. Use these keyword lists as candidates for outlier terms, or to identify themes associated with a cluster.
- Use the Charged Terms view in the Lexical Analysis tool (Figure 19) to look for keywords highly associated with lexicon terms (the keywords with a check-mark are already present in our lexicon, and other keywords in the list and are highly associated with the terms in our lexicon axis).

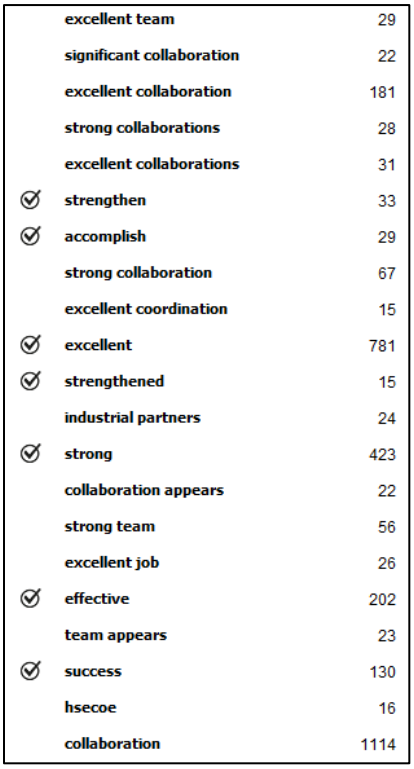

Figure 19. Charged Terms View within IN-SPIRE™

Our process would often start with selecting all the documents in the dataset to identify the top terms in the dataset. We would ask ourselves if we could explain the relationship between a given word and the axis and criterion being measured. We would outlier terms that appeared too generic or unrelated. We would perform this same process on each of the cluster centroids.

Galaxy cluster labels also provided good candidates for themes, but this often required some iteration with exploring terms and designating other terms as outliers. For some criterion, we found the cluster labels contained EERE office names, National Laboratory names, or names of other companies, organizations, or consortiums. Sometimes these were useful, but often we were striving to learn what feature or quality about these organizations were described in the record. Adjectives and verbs also occurred frequently in some clusters, and we would outlier these terms if we did not want IN-SPIRE™ to cluster comments around the use of these terms.

Cluster centroids allow us to inspect each individual cluster of documents that IN-SPIRE™ created for a dataset. Documents cluster together because they are mathematically similar to each other, but not necessarily because of a human understandable theme. Sometimes, documents cluster together because they share industry-specific terminology or other nomenclature. Sometimes IN-SPIRE™ will cluster documents together for mathematically

significant reasons that cannot be easily explained. We use the cluster centroids as hints to where clusters of thematically similar documents appear, but there are many false positives.

The assignment of individual documents to a cluster is determined based on the mathematical and statistical attributes of the documents in the dataset. Sometimes, documents are placed together because they use the same combinations of words but do not necessarily have the same meanings. Sometimes, two documents that mention the same concept are not necessarily clustered together because of other words in the document that also contribute to the mathematical signature of a document. The clustering algorithm of IN-SPIRE™ can only place documents in a single cluster. If a document contains individual thoughts, IN-SPIRE™ will assign the document to a single cluster, based on its overall mathematical similarity to other documents. Sometimes this cluster will contain documents that also contain one or more of the same individual thought concepts but sometimes it may not. Because of this, the IN-SPIRE™ clusters should be used as a navigational aid for locating potential clusters of similar content, and not as a precise tool for locating any and all content that is related to each other from a human comprehension standpoint. Our analysis often combined several IN-SPIRE™ clusters into a single theme, or expanded a cluster membership by including individual records that were spatially near a given cluster.

One point of caution with this approach, and something we were very conscious of during our analysis, was to avoid confirmation bias when examining a cluster. Confirmation bias is searching for information that matches our original thinking or hypothesis. By looking at keywords, we could hypothesize a topic area for a cluster, then simply seek out documents in that cluster that included this topic. Although we did use this process to confirm if our hypothesis was valid, we also took care to look at the documents within a cluster that did not match our hypothesized topic and questioned whether our hypothesis was correct, too narrow, too broad, or simply invalid. We acknowledge that some of our themes are very broad and could require further breakdown into subtopics and subthemes. This has been described more in the future work section (Section 7.4) of this document.

Charged Terms in the Lexical Analysis tool were hit and miss, but when they were a hit, they provided some good insights. One insight was identification of more subtle contexts around the use of terms (since charged entities tend to contain phrases in addition to single words). Another insight was a general validation of our lexicon. If the Charged Terms view showed us words that were intuitively related to the criterion and axis, we were more confident that our lexicon was measuring the concept accurately.

In all of these methods, we did not rely on keywords alone, but ultimately used the Document Viewer to read the comments themselves. Since IN-SPIRE™ was only analyzing records extracted from each comment, we wanted to make sure that we understood the larger context in which a record appeared within a comment. We would turn on highlighting for both search terms and lexicon terms. The visual cues in the Document Viewer enabled quick evaluation of the context of the lexicon term and any terms we searched for. Often, we would search for the keywords that described the criterion itself, as this would often pin-point to the portion of the original comment that was most relevant for our analysis.

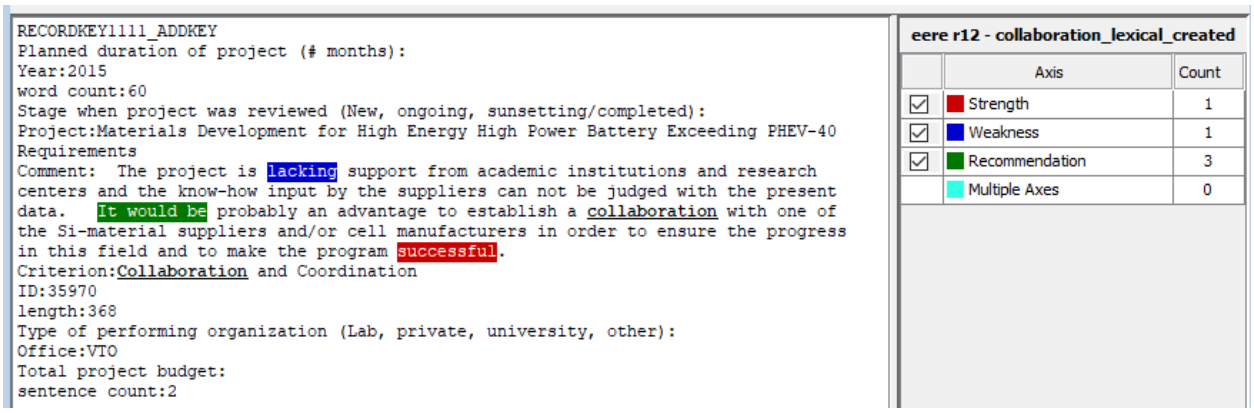

Figure 20. IN-SPIRE™ Document viewer highlighting search term "collaboration" and lexicon terms.

With this technique, the identification of significant keywords was data driven by IN-SPIRE™, but we used a human centric process of curating these keywords into themes by reading the actual comments themselves. Our goal was not to read every document, but to utilize IN-SPIRE™ to direct our reading towards highly populated clusters, so we could identify the prominent themes within each dataset. We would jump around to different regions of the galaxy, select cluster centroids, look at the top keywords, then read a few documents to see if we could determine a descriptive theme that associated these documents into a cluster. Once a cluster was located, we would add those documents to an IN-SPIRE™ group. We tried to find as many themes where documents did not belong to any of our curated theme groups as we could, by examining the cluster centroids in areas of the Galaxy, where documents did not belong to any of our curated theme groups.

For each dataset we analyzed, we provided a screenshot of the Galaxy view, showing the layout of where themes were located visually within each dataset. We used Groups within IN-SPIRE™ to organize documents and clusters into our curated themes, then assigned each Group a unique color. We manually added the labels and arrows to identify each theme and its location in the Galaxy.

We also used the IN-SPIRE™ Document Viewer to locate excerpts from comments to include in this report. After the positive identification of a theme, we manually scanned documents within the cluster to find positive examples of the theme within the comments. Although each cluster varied in size, we tried to find at least 4 comment excerpts, striving for 5 excerpts per theme.

Appendix B – Data Dictionary

We used the "DELETED PROJECTS" sheet within the "PPRD Peer Review Comments Final Database 2.23.21 w deletions 8-8-21.xlsx" Excel file. This sheet contained 70671 rows and 12 columns of data. Table 2 lists the columns from this sheet in the Excel file.

Table 2. Table of data columns from the comment spreadsheet used in our data analysis.

Column	Name	Description	Sample Value
A	ID	Number to uniquely identify each row	1
B	Office	EERE Office this project is associated with	AMO
C	Year	Year the review was performed	2018
D	Project	Name of the project being reviewed	Adsorptive Nitrogen Rejection from Natural Gas
E	Reviews for DELETED PROJECTS=1	Flag to indicate if this project is a non-R&D project to remove from the IN-SPIRE™ analysis. Rows with a 1 should be removed from IN-SPIRE™.	0
F	Add non-focus area Overview back in	Unsure of the meaning. This column was always blank	
G	Criterion	The criterion to which the review comment applies to	Accomplishments
H	Type of performing organization (laboratory, private, university, other)	This column was always blank. Future work would require populating this column for comparative analysis of different types	
I	Stage when project was reviewed (new, ongoing, sunsetting/completed)	This column was always blank. Future work would require populating this column for comparative analysis of different types	
J	Total project budget	This column was always blank. Future work would require populating this column for comparative analysis of different types	
K	Planned duration of project (# months)	This column was always blank. Future work would require populating this column for	

		comparative analysis of different types	
L	Comment	The actual comment as written by the reviewer	Excellent progress to date. It appears that all that could be accomplished in this sort of field test has been achieved, and the results are promising.

After PNNL split comments into separate records, we added the additional columns listed in Table 3 to the spreadsheet.

Table 3. Table showing additional columns added to the comment spreadsheet.

Column	Name	Description	Sample Value
M	PartialComment	An excerpt from the full comment that is found in Column L	Excellent progress to date
N	debug	An extra column to validate the splitting function used to break apart the full comment. Vertical pipe characters separate each PartialComment found in the full Comment	Excellent progress to date. It appears that all that could be accomplished in this sort of field test has been achieved, and the results are promising.

Appendix C – IN-SPIRE™ Dataset Creation

We used the Microsoft Excel/CSV option within IN-SPIRE™ to create our datasets. As described above, we created one primary dataset that imported the overall CSV file containing all partial comments, then used the subset function to create subset datasets for each criterion and concept we were measuring. This section describes how we created the initial dataset, using a CSV file that has been processed by our sentence splitter.

Step 1 – Create a new "Excel and CSV File" dataset

Step 2 – Add the .CSV file containing all of the reviewer comments

Step 3 – Create IN-SPIRE™ fields from the CSV columns as shown in Figure 21.

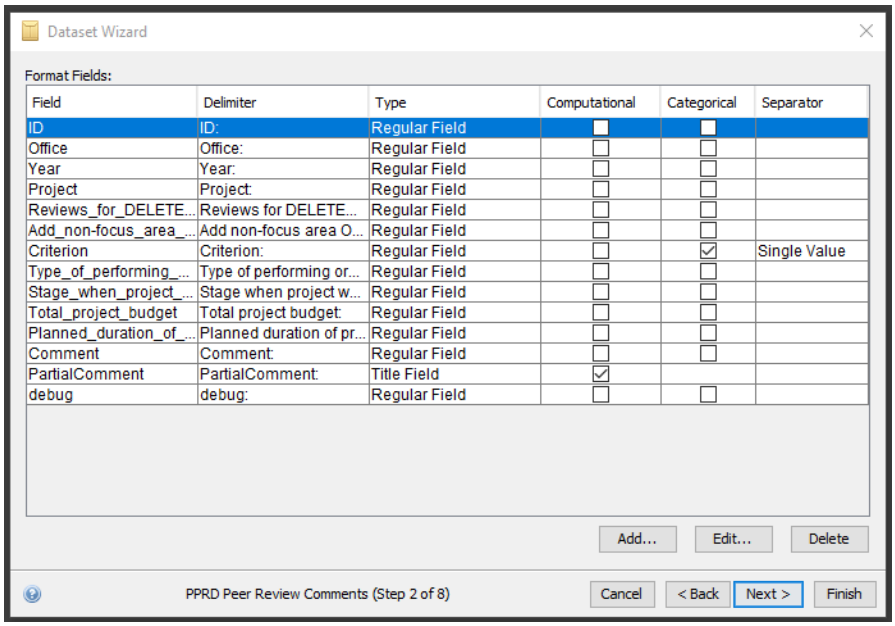

Figure 21. Dataset Field Specifications

We designated Criterion as a Categorical field, so we could use the Facets tool to easily subset our dataset by criterion. And we designated PartialComment as the only Computational field, to force IN-SPIRE™ to cluster content around the contents of this field. We also made this field our Title field for easier readability of comments in the Document Viewer. We did not mark any other field as Computational or Categorical. We also did not designate a Date field or make any fields be Number fields, as our spreadsheet did not contain any relevant number or date values.

Step 4 – No Entity Extraction

Step 5 – Default Stopwords

Step 6 – Load custom Stop Majors file. The Stop Majors file will exclude certain words from being considered as being statistically relevant, but IN-SPIRE™ will still full-text index these words to make them searchable.

Step 7 – No other custom dataset options, click Finish.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov