

Introducing Auto-Classification at a Major National Foundation

Text Analytics Forum, 11/16/21

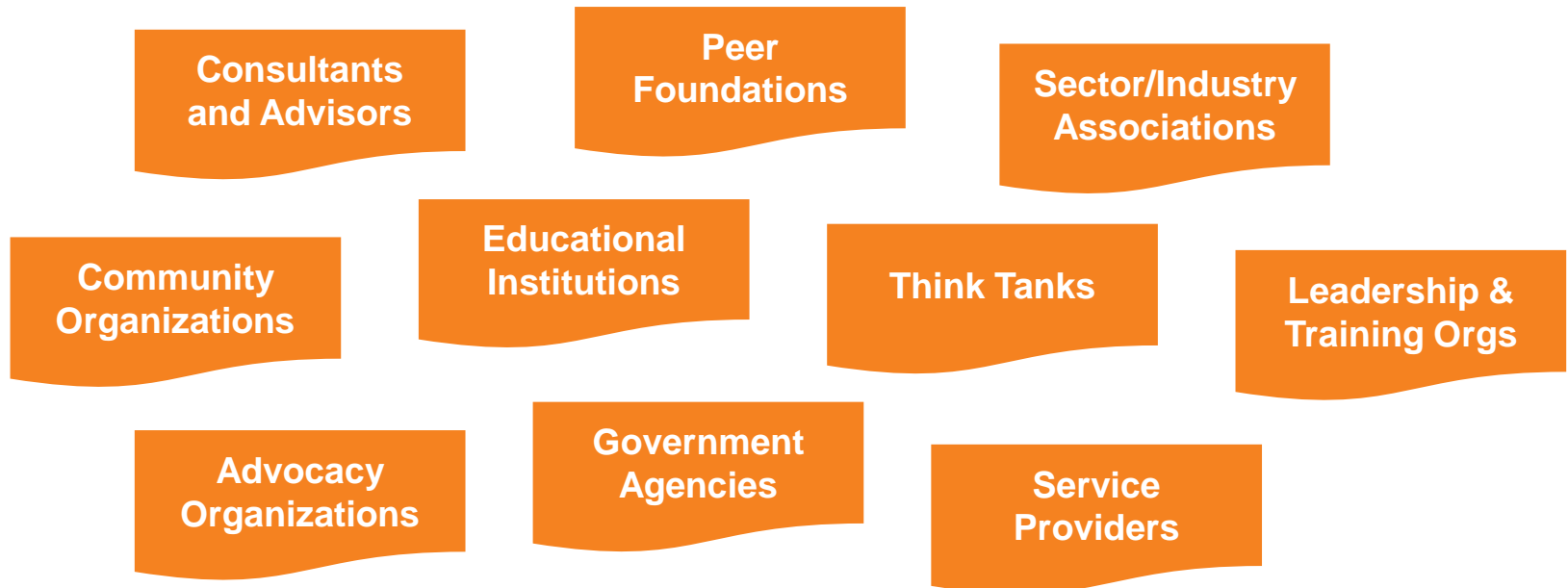
About RWJF

- Established in 1972, nation's largest private philanthropy dedicated solely to health
- Grounded by vision of a [Culture of Health](#) with health and racial equity at the center
- Aligned with vision, strategies in broad thematic areas
- In connection with grantmaking, a vast amount of knowledge exchange and creation with diverse collaborators nationwide



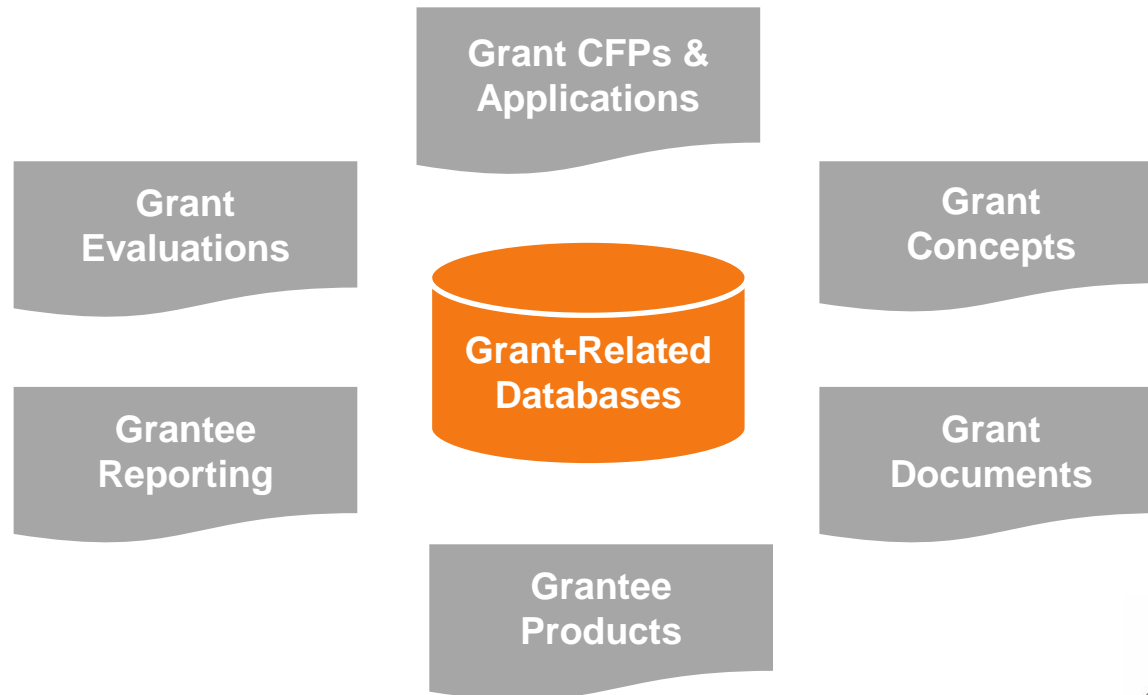
RWJF Knowledge Ecosystem: Collaborators

A sampling of the kinds of organizations we work with:



RWJF Knowledge Ecosystem: Grantmaking

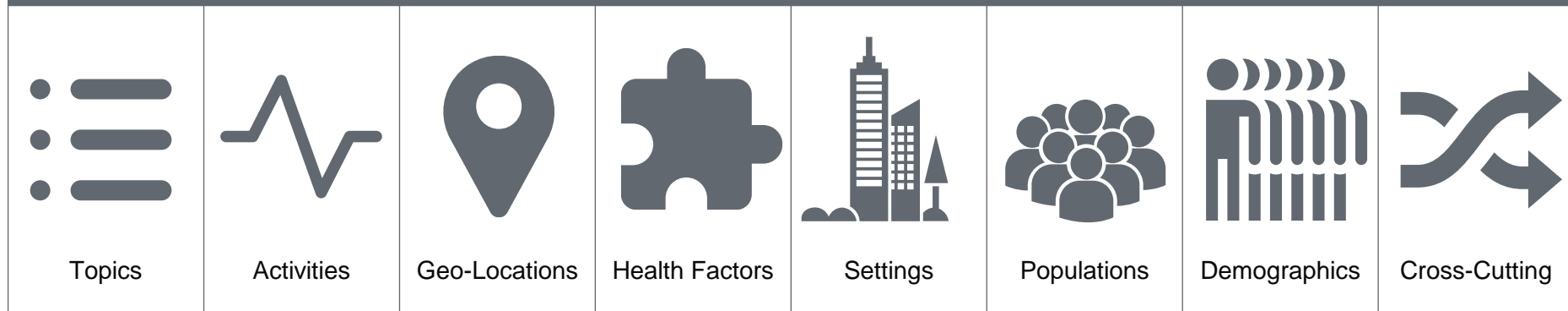
Through field relationships, we generate key knowledge and information through grant development and management



RWJF Knowledge Ecosystem: Taxonomy

- Established in 2019 as new enterprise-style terminology structure
- Current primary usage areas include Program Information Management System, Repository (grantee products), and public website (RWJF.org)
- Also strategic and integrated taxonomy management and governance

Current RWJF Taxonomy Model



RWJF Taxonomy Application (Current)

RWJF Taxonomy Definitions and Usage Guidelines

The RWJF taxonomy is a controlled vocabulary that supports classification of RWJF grants and related knowledge assets and presentation and retrieval of related data across RWJF information systems. If you have questions or feedback on taxonomy terminology or application, contact [Ari Kramer](#).

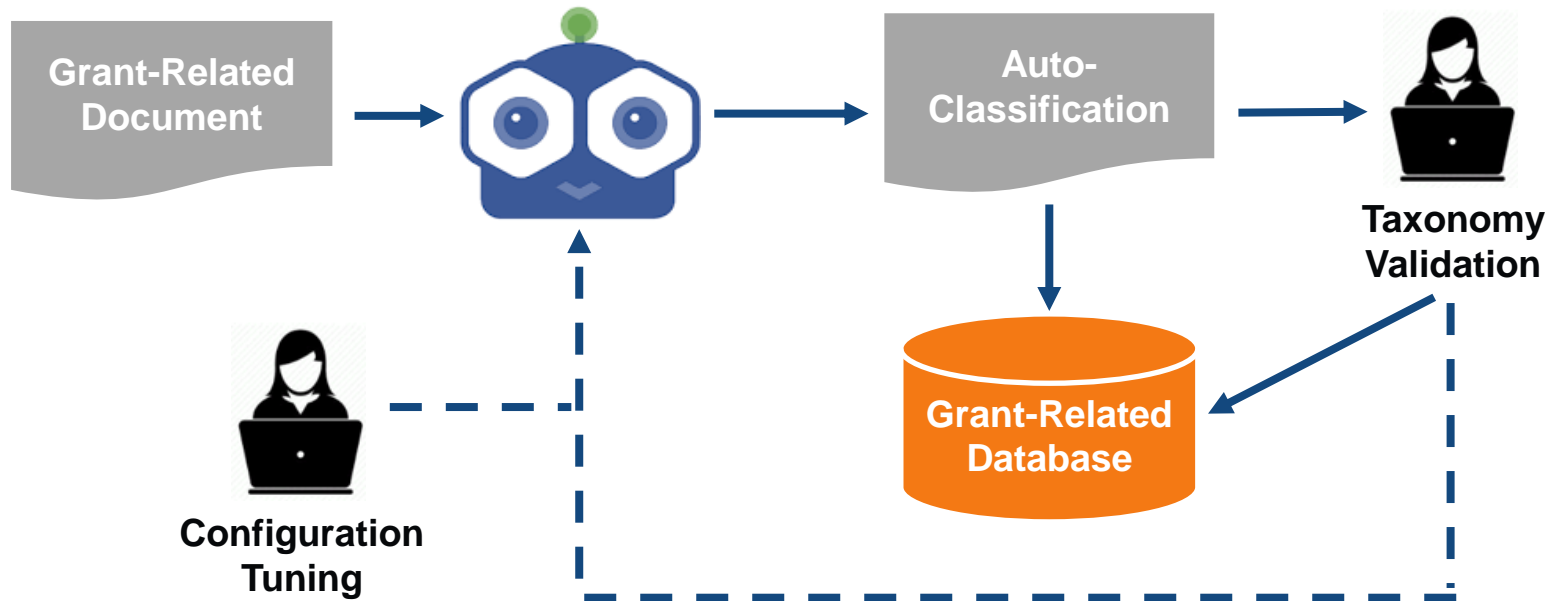
ACTIVITIES

Use to identify the major type of action RWJF is supporting to achieve the desired objective or the type of action a grant is primarily about. Most often there should be an activity, and ideally just one. Only add additional activities if a grant has more than one as a primary focus.

Advocacy	Efforts to inform public policy at the local, state, and federal levels; to bring particular issues or perspectives to the attention of policymakers; and/or to advance the interests of specific interest or constituent groups.
Capacity and Organizational Development	Efforts - such as such as staffing or management support, fund raising, or strategic planning to help institutions grow their ability to carry out their work. May also be used for general technical assistance and operating support.
Coalition and Network Building	Work to build alliances or systems of interconnected individuals, groups, or institutions for the purpose of elevating or advancing an issue or a relationship for a common interest and/or of achieving collective action.
Communications and Media	Content development, dissemination, or promotion to elevate and grow understanding on a particular issue and/or to foster the exchange of ideas among specific target audiences or the public.
Data and Technology	Projects focused primarily on data collection or sharing, or on technology or infrastructure to enable data collection or sharing. May also be used for other type of projects with major focus on technology or technology-related innovation.
Evaluation	A specialized form of research to determine the extent to which a grant or program achieves its targeted objectives or to assess the effectiveness or impact of other RWJF-supported activities.



RWJF Taxonomy Application (Future)



Text Analytics at RWJF: Background

Next step in RWJF taxonomy application – text analytics

Initial trial – reinforced importance of finding right software

- The product RWJF used worked great for short text, but not the longer content.

Focus of RWJF's 2020-2021 engagement w/ KAPS Group

- **Mini-POC**: A 'real-world' prototype which demonstrates text analytics value and provides foundation for development
- **Software Evaluation**: Evaluate and compare multiple products, select a winner, complete initial implementation



Text Analytics at RWJF: Mini-POC

Selection of 10 Categories – New Taxonomy

- 6 Topics, 3 Populations, 1 Activity

Selection of 20 documents per category

- Good examples of documents about category

Develop content structure model rules

- Summary / Introduction and other

Develop categorization rules for each category – 90%+



RWJF Mini-POC: Category Selection

Task : Select categories from taxonomy and documents for each category

Activities	Topics	Populations
<ul style="list-style-type: none"> ○ Advocacy ○ Capacity & Organizational Development ○ Coalition & Network Building ○ Communications & Media ○ Data & Technology ○ Evaluation ○ Idea Sourcing ○ Leadership, Education & Training ○ Pilot Project ○ Research & Analysis 	<ul style="list-style-type: none"> ○ Child & Family Well-Being ○ Childhood Obesity ○ Disease Prevention & Health Promotion ○ Early Childhood ○ Health Care Coverage & Access ○ Health Care Quality & Value ○ Health Disparities ○ Health Leadership Development ○ Nurses & Nursing ○ Public & Community Health ○ Social Determinants of Health 	<ul style="list-style-type: none"> ○ Activist/Advocate ○ Business (General) ○ Caregiver ○ Faith-Based ○ Government ○ Health Professional ○ Immigrant or Migrant ○ Inmate/Former Inmate ○ LGBTQ Person ○ Non-Health Professional ○ Nonprofit/Philanthropy ○ Parent ○ Person with Disability ○ Person with Disease or Illness ○ Policymaker ○ Student



RWJF Mini-POC: Content Structure Model

- Develop a taxonomy of document types and structures
- Identify key sections for each document type
 - Abstract – human judgement of aboutness of document
 - Negative sections to exclude
- Identify text indicators for each document section
 - Abstract, Executive Summary, Introduction, etc.
- Consider issues such as variability of writing, range of text indicators, documents without clear sections, etc.



RWJF Mini-POC: Content Structure Model

Providing actuarial analyses and modeling of health reform ideas to stabilize individual insurance markets and continuing RWJF's actuarial challenge

Fund Description: To continue dissemination and analytical activities associated with the results of the 2017 RWJF Actuarial Challenge, in which teams of actuaries proposed solutions to stabilize the individual insurance market.

SUMMARY

This project will continue the work of the RWJF Actuarial Challenge. The Actuarial Challenge took place in early 2017. The final results included policy suggestions for stabilizing the individual market, including elements such as reinsurance, auto enrollment, and other market reforms. Milliman organized the challenge and simulated the winning proposals, providing estimates of how they would impact enrollment and public and private spending. As the prospect for bipartisan health reform increases, there is an increased demand for disseminating these results and for potentially engaging in some additional actuarial modelling. The challenge process and results are reviewed as technically credible and politically nonpartisan. As 1



RWJF Mini-POC: Content Structure Model

Providing actuarial analyses and modeling of health reform ideas to stabilize individual insurance markets and continuing RWJF's actuarial challenge

Fund Description: To continue dissemination and analytical activities associated with the results of the 2017 RWJF Actuarial Challenge, in which teams of actuaries proposed solutions to stabilize the individual insurance market.

SUMMARY

This project will continue the work of the RWJF Actuarial Challenge. The Actuarial Challenge took place in early 2017. The final results included policy suggestions for stabilizing the individual market, including elements such as reinsurance, auto enrollment, and other market reforms. Milliman organized the challenge and simulated the winning proposals, providing estimates of how they would impact enrollment and public and private spending. As the prospect for bipartisan health reform increases, there is an increased demand for disseminating these results and for potentially engaging in some additional actuarial modelling. The challenge process and results are reviewed as technically credible and politically nonpartisan. As t



RWJF Mini-POC: Content Structure Model

Providing actuarial analyses and modeling of health reform ideas to stabilize individual insurance markets and continuing RWJF's actuarial challenge

Fund Description: To continue dissemination and analytical activities associated with the results of the 2017 RWJF Actuarial Challenge, in which teams of actuaries proposed solutions to stabilize the individual insurance market.

SUMMARY

This project will continue the work of the RWJF Actuarial Challenge. The Actuarial Challenge took place in early 2017. The final results included policy suggestions for stabilizing the individual market, including elements such as reinsurance, auto enrollment, and other market reforms. Milliman organized the challenge and simulated the winning proposals, providing estimates of how they would impact enrollment and public and private spending. As the prospect for bipartisan health reform increases, there is an increased demand for disseminating these results and for potentially engaging in some additional actuarial modelling. The challenge process and results are reviewed as technically credible and politically nonpartisan. As th



RWJF Mini-POC: Content Structure Model

Providing actuarial analyses and modeling of health reform ideas to stabilize individual insurance markets and continuing RWJF's actuarial challenge

Fund Description: To continue dissemination and analytical activities associated with the results of the 2017 RWJF Actuarial Challenge, in which teams of actuaries proposed solutions to stabilize the individual insurance market.

SUMMARY

This project will continue the work of the RWJF Actuarial Challenge. The Actuarial Challenge took place in early 2017. The final results included policy suggestions for stabilizing the individual market, including elements such as reinsurance, auto enrollment, and other market reforms. Milliman organized the challenge and simulated the winning proposals, providing estimates of how they would impact enrollment and public and private spending. As the prospect for bipartisan health reform increases, there is an increased demand for disseminating these results and for potentially engaging in some additional actuarial modelling. The challenge process and results are reviewed as technically credible and politically nonpartisan. As



RWJF Mini-POC: Content Structure Model

Providing actuarial analyses and modeling of health reform ideas to stabilize individual insurance markets and continuing RWJF's actuarial challenge

Fund Description: To continue dissemination and analytical activities associated with the results of the 2017 RWJF Actuarial Challenge, in which teams of actuaries proposed solutions to stabilize the individual insurance market.

SUMMARY

This project will continue the work of the RWJF Actuarial Challenge. The Actuarial Challenge took place in early 2017. The final results included policy suggestions for stabilizing the individual market, including elements such as reinsurance, auto enrollment, and other market reforms. Milliman organized the challenge and simulated the winning proposals, providing estimates of how they would impact enrollment and public and private spending. As the prospect for bipartisan health reform increases, there is an increased demand for disseminating these results and for potentially engaging in some additional actuarial modelling. The challenge process and results are reviewed as technically credible and politically nonpartisan. As the



RWJF Mini-POC: Content Structure Rules

Count terms in list and first 100 words unless negative terms within 7 words

Count terms in list and within 500 words after a Document Summary Indicator unless negative terms* within 7 words

- Document Summary Indicators – 29 terms “Executive Summary”, “Issue Brief”, “Abstract”

** Negative terms are terms that often show up but do not match the focus of a category – e.g., “Obesity” in “Child & Family Well-Being”*



RWJF Mini-POC: Average Scores

	Recall	Precision	Precision Top 10
With Sections	95%	92%	99%
Full Text	71%	41%	81%
Difference	24%	51%	18%



RWJF Mini-POC: 4+ Topic Project

Problem: Grants with over-indexing

- Over 8,000 grants with 4 or more Topics (35% of funding items)
- Excessive classification makes grant info harder to find and use

Solution: Reclassify with text analytics

- Complete development of auto-categorization rules for Topics facet, and apply to 8000 over-indexed grants
 - Key Issue: accuracy scoring

Results: Provided better grant data and ROI

- Leveraged POC to increase the quality of critical RWJF data, and demonstrate the value of text analytics investment



RWJF Text Analytics Software Evaluation

Stage 1 – User Research

- High-level research – RWJF information environment
- Requirements – user interviews

Stage 2 – Vendor Research

- Pass 1 – range of options – best fit for RWJF
- Pass 2 - selection process – **45 to 5** vendor candidates
- 5 Demos – deep dive

Stage 3 – POC and Selection

- POC top 2 candidates

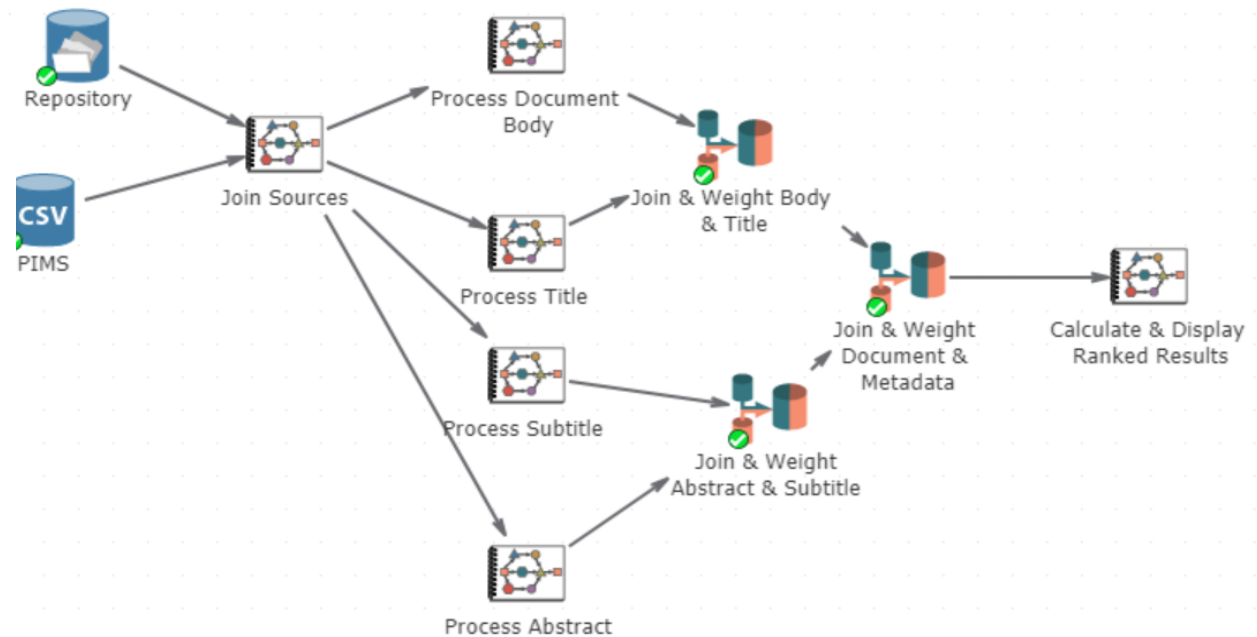


RWJF Text Analytics Software Evaluation

Selection: [Megaputer Polyanalyst](#)

Rationale: Overall best fit for RWJF environment, needs and budget

- Text Analysis
 - Abbreviation Expansion
 - Auto Taxonomy
 - Distinct Texts
 - Emerging Trends
 - Entity Anonymization
 - Entity Extraction
 - Fact Extraction
 - Fragment Analysis
 - Grammar Check
 - Index
 - Join Taxonomies
 - Keyword Extraction
 - Language Detection
 - Link Terms
 - Reconstruct Entities
 - Search Query
 - Sentiment Analysis
 - Spell Check



Text Analytics at RWJF: Looking Ahead

- Initial implementation – Repository (grantee products)
- Complete configuration for rest of RWJF taxonomy
- Present outcomes to key internal stakeholder groups
- Prepare for next phase of implementation (PIMS/grants)
- Conduct more research about additional use cases
 - Auto-categorization, Entity-Fact Extraction, New?
- Planning for long-term management & resource allocation



Key Takeaways and Closing Thoughts

- Auto-categorization is a key foundation
- Building strong working relationships is critical!
- A great project manager is a wonderful thing
- Be thoughtful and systematic on software selection
- Stay aligned with related organizational strategy areas
- Educate and engage key stakeholders along the way
- Don't go faster than needed – move ahead at a smart pace

