INSTITUTE FOR DEFENSE ANALYSES

Getting out of the Corner: Tapping into Unstructured Data for Improved Organizational Analyses

Margaret D.M. Barber
Terri J. Walsh
Anthony L. Johnson
Christophe L. McCray

September 2021
Approved for public release;
distribution is unlimited.
IDA Document NS D-21568
Log: H 21-000054

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
The work was conducted by the Institute for Defense Analyses (IDA) under CRP C6581.

For More Information:
Ms. Terri J. Walsh, Project Leader
twalsh@ida.org, 703-845-2468
ADM John C. Harvey, Jr., USN (ret) Director, SFRD
jharvey@ida.org, 703-575-4530

Getting out of the Corner: Tapping into Unstructured Data for Improved Organizational Analyses

Introduction

Reorganization within a government agency can be the result of a single or a combination of drivers ranging from new legal mandates, to the need to reduce spending, to the desire to support strategic priorities, to name a few. In large agencies like the Department of Defense (DoD), even minor organizational changes can lead to unintended consequences that can harm the Department's ability to manage its business. Traditionally, a small set of tools that draw on some combination of personal experience, reason, and block-and-line diagrams is used to assess the need for organizational change and avenues for improvement. But the complexity of the Office of the Secretary of Defense's (OSD) organizational structure and its rate of change argue for more rigorous approaches, particularly as the Department wrestles with the possibility of a new round of reorganization.

This article proposes an approach—the Barber-Walsh Method for Policy Analysis—that leverages organizational data, most of which is unstructured, to identify gaps and overlaps across similar sets of roles and responsibilities. The method we propose moves the Department away from anecdotes and case studies and towards sets of evidence-based recommendations as to how offices within OSD can more effectively execute their legal requirements.

To illustrate how we have applied this method, we use an example of how DoD implemented new legal requirements for security cooperation programs and activities levied in the Fiscal Year 2017 National Defense Authorization Act (FY17 NDAA).[1] Because security cooperation involves so many actors across DoD, implementing FY17 NDAA's requirements meant sweeping changes needed to be made to DoD's security cooperation processes and practices. Policy, namely Department of Defense Issuances, capture how the Department implements law. We describe a methodology that we used to assess whether organizations charged with security cooperation oversight had implemented

[1] The National Defense Authorization Act for Fiscal Year 2017, Public Law 114-328, U.S. Statues at Large 130 (2016): 2000-2968.

the FY17 NDAA requirements. At the request of the organizations involved, specific names and NDAA requirements are unnamed.

We begin by outlining how we propose to leverage an organization's unstructured data to improve how policy implements law. Next, we use the FY17 NDAA example to illustrate how the methodology was applied. This method has been used against other questions of strategy and policy, so it need not be restricted in its application.

What is unstructured data and how do defense analysts deal with it?

For the purposes of this article, *unstructured data* refers to data that are not stored in a traditional database and are therefore not well organized for analysis. Such data lack a data model and related rules and organizational principles. But unstructured data is growing—around 55–65% annually—and represents the majority of data any organization has. It therefore cannot be omitted from organizational and policy analyses for an organization as complex as the Department of Defense.[2]

One of the most common approaches to harnessing unstructured data has been text analytics and artificial intelligence technologies.[3] Text analytics may help to analyze data in a text format—like policy and guidance documents—but it can be insufficient or impractical. For example, DoD might block or restrict the use of text analytics tools on some systems for security reasons. Mining and organizing unstructured data can be labor-intensive and time-consuming, which often leads to using anecdotes more readily to hand. Unfortunately, anecdotes tend to be either subject to selection bias because they tell a good news story, or few in number chosen for their availability rather than suitability. Harnessing unstructured data within a structured format helps analysts understand what the data look like in total so that they can either analyze these data themselves or better select cases for more robust analyses to support key decision-makers.

A New Process for Structuring Unstructured Data

Our approach comprises identifying and refining requirements, verification, and validation, with corresponding steps for each. While it is depicted below as cyclical, in

---

[2] Bernard Marr, "What Is Unstructured Data and Why Is It So Important to Businesses? An Easy Explanation for Anyone," *Forbes* (16 Oct 2019); "What Is Unstructured Data?" NetApp, n.d.

[3] See, for example Marr, "What Is Unstructured Data and Why Is It So Important to Businesses…"

application, analysts often find they need to repeat sets of steps to refine and improve the way in which unstructured data are captured.

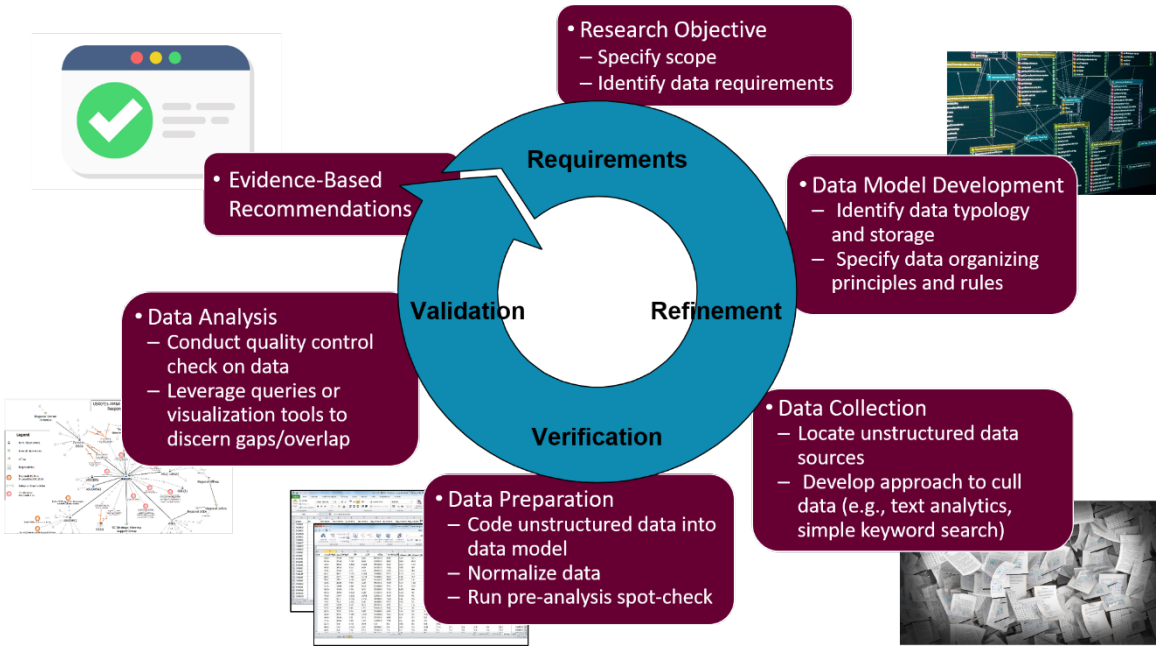

Figure 1. Approach to Leveraging Unstructured Data for Organizational Analyses

Research Objective and Data Model Development

In our approach, the objective and scope of the research drive data requirements—they define the data needed to do the analyses and they specify the level of detail at which it must be collected. Based on the scope, the analyst forms a data model, perhaps the most critical step. The model must incorporate two elements: (1) basic data that analysts can use to characterize the analytical space, and (2) more specific data that enable analysts to identify and define critical metrics (captured in fields). For the data model to be useful, it requires rules for entering data. Rules should guide data input for each field and specify relationships between data; they also must be captured in a format that is sharable across the team of coders. Establishing these rules early reduces the error rate and creates opportunities for quality control later in the process. Creating the data model should inform the next step, which is data collection.

Data Collection

Pulling relevant data out of unstructured sources to enter into a structured data model begins with identifying where relevant data are and designing approaches to mine it. Unstructured data can reside in many formats—numerical tables, graphical representations, timelines and schedules, budget figures to briefings, policies, guidance,

and directives. Unstructured data can also reside in the minds of experts, which can only be captured via structured interviews. The process and techniques for extracting data must anticipate and accommodate these various formats and types. Analysts can use the data model they created in the first step to identify keywords and to zero in on the information most essential to the analyses; they can then use those terms in their document and media searches, regardless of the data source or format. Analysts can also use the model to develop and design interview scripts or questions to draw out the most important data from an interviewee. Whatever the collection strategy, the goal of collection must be transferring unstructured data into the structured data model.

Data Preparation

In order for unstructured data to be useful, it must be made suitable for a more structured data model during the *Data Preparation* step. Data preparation consists of multiple tasks, including coding data, normalizing data, and defining terms. It is critical during this step that analysts adhere to pre-established and disciplined data collection and input protocols, and is often useful to test these protocols by coding sample data. This helps identify coding discrepancies, points to clarify, and issues with data granularity.[4] The more analysts work upfront to prepare the data for entering into a structured format, the less error and discrepancy later. Data preparation often informs refinements to the data model and data collection steps, so these steps may be repeated several times to ensure the data are well-organized and the model appropriately reflects the analytical requirements of the work.

Data Analysis and Developing Evidence-Based Recommendations

The final two steps of this methodology are *Data Analysis* and *Recommendations*. Data analysis begins with quality control. An ideal approach to quality control is to assign two data analysts to code the data—one to code each data point. The first analyst does a complete coding of each data source—e.g., a document, interview notes. The second analyst also fully codes the data for the same source. When complete, both analysts review their coding together to identify points of disagreement and then work through and resolve those disagreements; this ensures that the coders maintain a common inter-

[4] Data visualization is another useful tool for spot-checking the quality of the data model and the newly entered data. Visualization tests the data model to ensure that relationships between data points have been accurately and comprehensively captured. Visualization can also identify where data needs additional normalization to ensure they are not mis-labeled or mis-interpreted in ways that impact the analytical effort and skew the research results.

pretation of coding rules. The process is highly collaborative and requires intensively exploring and discussing coding disagreements so that the reasons for how and why they were resolved can be communicated to the entire team. Not uncommonly, research teams have neither the time, the resources, nor the need to do this level of quality control. In such cases, the second analyst assigned the coder role can conduct a randomized coding spot-check.

After quality control is complete, the analysts can review the data for findings. The first, often most straight-forward, step in the analysis phase is to count. Counting data points in different fields within the data model helps generate a baseline understanding of the analytical space—it identifies prevalent characteristics, omissions, and skew. After characterizing the space, the analyst can then select from a variety of analytic tools—from coded queries to pivot tables to statistical analyses to further visualization. Ultimately, it will be up to the analyst to choose the most appropriate technique to answer the research question. However, by this point the analyst should have a more complete dataset from which to draw conclusions.

Benefits of Using the Unstructured Data Approach

We advocate for applying a structured framework for storing and analyzing unstructured data. During the past two years, our research team used this approach to do policy analyses of two research questions: (1) Has DoD met specific legal requirements related to security cooperation programs and activities? and (2) Does DoD policy provide sufficient implementation guidance to organizations responsible for meeting those requirements? To answer those questions, the research team designed a data model that enabled analysts to assess relevant roles and responsibilities for substance and clarity, and to make judgments about the quality of the Department's implementation efforts. This particular data model relied on single actor-task pairings to determine how legal requirements were translated into assignments of key tasks across the Department; which DoD Components were affected by the policy, and what coordination and cooperation is required for implementation; the specific authorities and tasks the Department's implementation policy assigns to DoD officials; and the categories that tasks fall into to illustrate where the bulk of implementation lies. Together, those aspects of the data model reveal a clear picture of who does what and how responsibilities have been delegated to

implement the law. This allowed us to discern whether formally delegated responsibilities implement all aspects of the law and are clear enough to be acted upon in order to accomplish the law's intent.[5]

Two sources of data were necessary to answer the research questions: the DoD policymaker expertise and DoD policy and guidance documents. We began collecting data by reviewing all DoD policy (defined narrowly for this effort as DoD Issuances) to better distinguish it from guidance documents, which we defined as any prescriptive document that was not a formal DoD product. That distinction enabled us to specify which information would be best for DoD policy (as specified in Table 2 of DoDI 5025.01), and which would be better served in guidance documents. We conducted a keyword search of all DoD Issuances and pulled more than 90 such documents for the analysis. We then generated a list of relevant stakeholders and developed a standardized script for interviews so that data was consistently collected, and the process was repeatable—small teams of analysts interviewed members of the stakeholder community using a common set of questions and data capture forms. We noted what we termed "blind spots"—areas where stakeholders were unavailable—and therefore for which data could not be collected. Ultimately, the interviews allowed us to compare what practitioners perceived their responsibilities to be, relative to the new requirements in the FY17 NDAA, with what DoD issuances proscribed.

Once we collected the data, we then began to input it into our data model. Because our data model relied on specific actor-task pairs as the unit of analysis, this often meant that a single responsibility in an issuance could yield multiple rows of data, because responsibilities are usually assigned in related clusters to clusters of actors. This level of granularity was critical to discerning whether policy assignments were clear and comprehensive. Combined, this research yielded a dataset with approximately 2,000 records. We paired data coders for quality control and resolved disputes by consensus. Once quality control was complete, we imported the data into IBM's Analyst Notebook, a visualization tool that served as a second round of verifying and validating the data and as an analytic tool as well.

Once data normalization was complete, we began the analysis by characterizing the distribution of relevant roles and responsibilities across the Department, using the categories captured in the data model in a single field in which the analyst could select the

[5] See for example, U.S Department of Defense, Office of the Chief Management Officer, DoD Instruction 5025.01: DoD Issuances Program, (Washington, 2019).

relevant category from a drop-down box. This allowed us to map where the Department housed different responsibilities that were important to executing the law, so we could then assess whether this was a comprehensive approach, both in terms of the law itself and DoD's issuances standards. We also used visualization (Figure 2) to comprehend how many informal tasks related to new FY17 NDAA legal requirements were concentrated within specific organizations in order to determine whether formally assigned responsibilities sufficiently covered them. In the case of this map, formal (assigned) responsibilities are depicted in blue, and informal (executed) are in red. Responsibilities in purple indicate alignment between assigned and executed tasks.

The message of the map is clear—there is little alignment between what is written in formal DoD issuances and the tasks that are actually executed to implement new legal requirements. This does not necessarily mean that policy must be changed; rather, it flags for the analyst the potential for gaps and overlap when interpreting how law should be carried out.

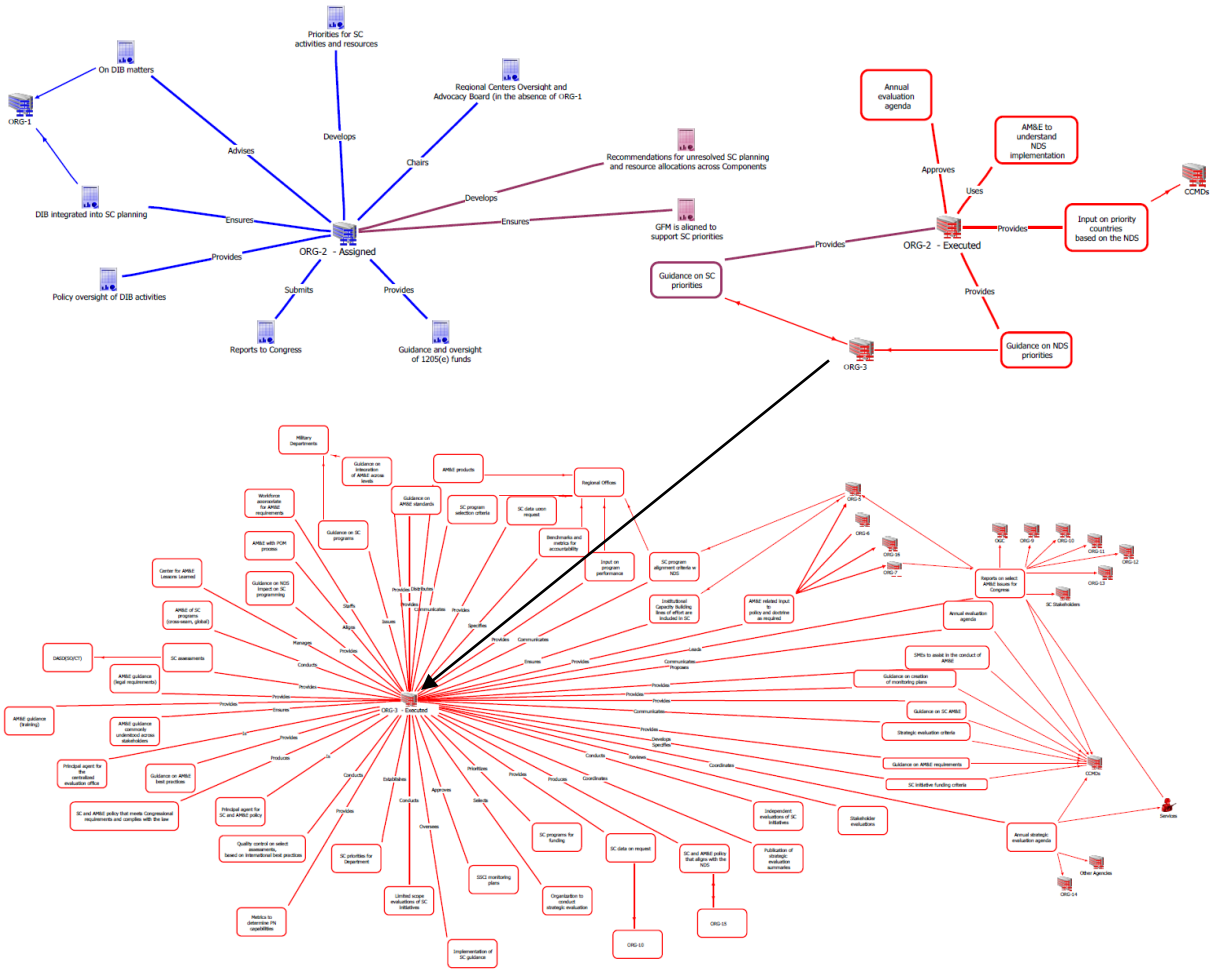

Figure 2. Visualizing Organizational Roles and Responsibilities

Mapping the roles and responsibilities also enabled us to identify overlap, policy areas that needed to be clarified, and gaps across responsibilities. The result of the analysis was a set of recommendations to streamline and improve the clarity of existing policy to implement the legal requirements in the FY17 NDAA.

Conclusion

We have offered the reader a method to improve organizational and policy analyses by harnessing unstructured data within an organization. The Barber-Walsh Method implements a logical, transparent, and reliable analytical process that enables policy analysts to develop a comprehensive picture of any policy and strategy environment they want to understand better. Our analytical method facilitates collaboration, and relies upon a logical framework that is easily explained to decision-makers. We understand variations of this approach are sure to have already been applied to similar questions.

Our purpose here is to offer a tested, validated, and established process that fully leverages the unstructured data across DoD and enables more robust recommendations to inform organizational change.

———

The authors would like to thank Michael Fitzsimmons, Clark Frye, Wade Hinkle, Russell A. Keller, Allison L. King, Aaron C. Taliaferro, and Jennifer M. Taylor for their advice and contribution to this work.

Dr. Margaret D.M. Barber is a research staff member at the Institute for Defense Analyses in the Joint Advanced Warfighting Division. She has spent the past decade developing unstructured data approaches to strategy and policy analyses supporting the Departments of Defense and Homeland Security.

Terri J. Walsh is a research staff member at the Institute for Defense Analyses in the Strategy, Forces And Resources Division. She…

Anthony L. Johnson a research staff member at the Institute for Defense Analyses in the Joint Advanced Warfighting Division. He…

Dr. Christophe L. McCray is a former IDA Research Staff Member who has worked for over 20 years in Science, Technology and National Security.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From – To)
XX-09-2021	Final	

4. TITLE AND SUBTITLE	5a. CONTRACT NO.
Getting out of the Corner: Tapping into Unstructured Data for Improved Organizational Analyses	5b. GRANT NO.
	5c. PROGRAM ELEMENT NO(S).

6. AUTHOR(S)	5d. PROJECT NO.
Margaret D.M. Barber Terri J. Walsh Anthony L. Johnson Christophe L. McCray	5e. TASK NO.
	5f. WORK UNIT NO.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NO.
Institute for Defense Analyses Strategy, Forces and Resources Division 4850 Mark Center Drive Alexandria, VA 22311-1882	IDA Document NS D-21568 Log: H 21-000054

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR'S / MONITOR'S ACRONYM(S)
Institute for Defense Analyses Strategy, Forces and Resources Division 4850 Mark Center Drive Alexandria, VA 22311-1882	11. SPONSOR'S / MONITOR'S REPORT NO(S).

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Reorganization within large agencies like the Department of Defense can lead to unintended consequences that can harm the organization's ability to manage its business. Traditional tools for reorganization are no longer sufficient to account for the complexity of the Office of the Secretary of Defense's (OSD) bureaucracy and its rate of change. This article proposes an approach—the Barber-Walsh Method for Policy Analysis—that leverages unstructured organizational data to identify gaps and overlaps across roles and responsibilities. The method we propose moves the Department away from anecdotes and case studies and towards sets of evidence-based recommendations as to how offices within OSD can more effectively execute their legal requirements.

15. SUBJECT TERMS

unstructured data, policy analyses, security cooperation, roles and responsibilities, organizational change, FY17 NDAA

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NO. OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ms. Terri J. Walsh
U	U	U	UU	14	19b. TELEPHONE NUMBER (Include Area Code) 703-845-2468

This page is intentionally blank.