FEMA

Department of Homeland Security
Federal Emergency Management Agency

Modeling and Data Working Group: Recovery Phase Analysis

DRAFT

August, 2016

Contents

Introduction

The Modeling and Data Working Group (MDWG), chaired by the Director of FEMA's Planning and Exercise Division, Response Directorate, was appointed by the Emergency Support Function Leadership Group (ESFLG) to identify and characterize the models and datasets that are used across the federal interagency in support of operational decision making for emergency management. MDWG members include subject matter experts, program managers, and program directors representing each of the Federal Emergency Support Functions (ESFs). The MDWG has produced the ESFLG Model and Data Inventory (MoDI),[1] an interactive, web-based resource that collates information about models and datasets in use across the Federal emergency management community for earthquakes, hurricanes, improvised nuclear device (IND), flood, and, as of this report, biological hazard scenarios.

The MoDI includes models and datasets that support all phases of emergency management and this report outlines an effort to identify those datasets and models that are used specifically to support recovery efforts. Both the ESFLG and the Recovery Support Function Leadership Group (RSFLG) were engaged to identify agencies, divisions, and groups to interview for this project. Based on interviews across the interagency emergency management community with senior leadership, program leads, and subject matter experts, the datasets and models used for recovery efforts were identified and assessed by how they are used to support recovery.

The datasets and models used to support recovery efforts across the Federal interagency were identified through previous MDWG efforts in developing the MoDI and by a series of 34 additional interviews conducted with 59 people representing 16 federal agencies, divisions, or groups (see Appendix A). Interviewees were identified through recommendation by the RSFLG and subsequent further recommendations by those interviewees.

This report contains three parts: a brief background of the recovery phase and information used to support recovery in relation to other phases of emergency management, discussion of quantitative and network analysis results, and identification of recommendations to better integrate recovery phase focused tools into the overall information network supporting emergency management.

Background: Recovery Phase

The response phase of emergency management is focused on life-saving and life-sustaining activities, stabilizing the incident, and protecting property and the environment. Like response operations, recovery phase activities begin early in the event, but typically continue much longer and can extend for years after the event. Thus, recovery overlaps with response, but also continues well beyond to support sustained efforts to meet housing needs, provide for health care (including emotional well-being), restore infrastructure and the local economic base, and initiate planning to mitigate the consequence of future disasters (Figure 1).

The first major question in recovery support operations is whether the event is large enough that community resources will be overwhelmed and a Federal response and interagency coordination will be required. Subsequently, Federal assistance programs need to be identified for activation and

[1] As of the publication date of this report, the MoDI is accessible at: http://gis.fema.gov/Model-and-Data-Inventory/

communities identified for support. This information is anchored in impact estimates and assessments of population and community vulnerability. Data requirements also include estimates for debris removal and road conditions to identify and resolve barriers to deploying personnel and materials to support recovery operations. These data requirements overlap with many of the event-specific data requirements for response, such as the need to understand what area was impacted, the severity of damage, and the population and infrastructure in the impacted area. Recovery also requires additional detailed information that is not used in response. Recovery operations require a more complete understanding of the financial needs of local businesses, the value of the property belonging to individual homeowners, whether that property was insured, and details about the health and financial stability of those impacted. As a result, recovery information often includes sensitive personal information that must be protected.

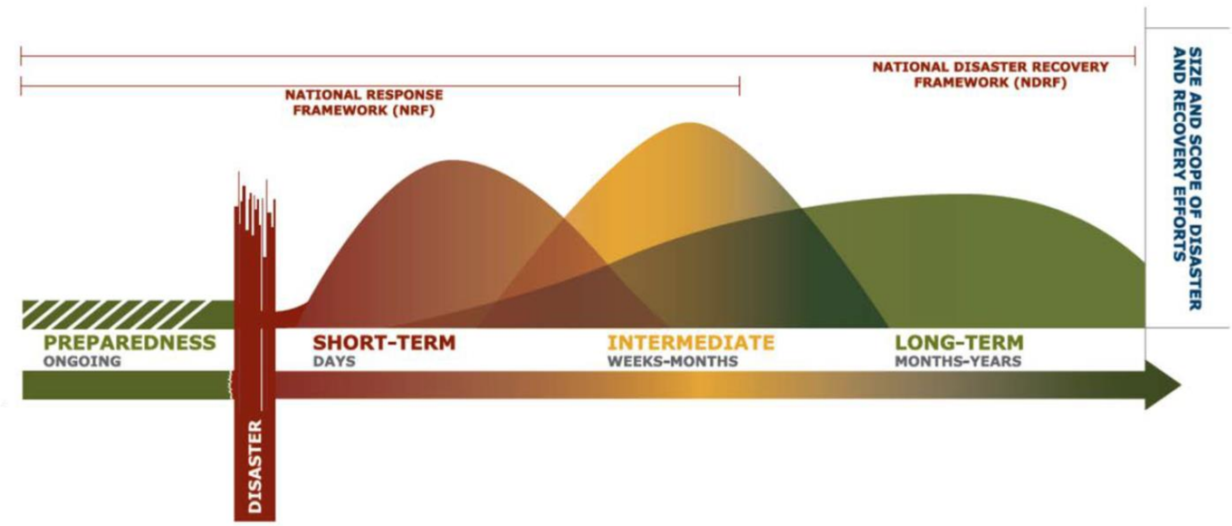

Figure 1. Timeline of the Recovery Phase, as described by the National Response Framework and National Disaster Recovery Framework. This timeline illustrates the continuum of short-term, intermediate, and long-term phases of recovery following a disaster and demonstrates the overlap between the National Response Framework and the National Disaster Recovery Framework pre-event and in the days immediately following a disaster. Figure reproduced from the National Disaster Recovery Framework.[2]

Recovery is driven by actions at the local level with Federal resources supporting recovery efforts for large-scale events. Therefore, much of the data and analysis performed is limited to the specific communities affected and are often not applicable to other geographic regions. The MoDI includes only those datasets or models used to support Federal emergency management decision making, so does not include datasets strictly available or relevant to State and local recovery.

[2] U.S. Department of Homeland Security (2016). *National Disaster Recovery Framework.* 2nd Ed. Retrieved from http://www.fema.gov/media-library/assets/documents/117794. Accessed 24 Aug. 2016.

Results

Results Overview

- Recovery tools are not well integrated with datasets and models used to support other phases or emergency management
- Response tools outnumber recovery tools for every hazard in the MoDI
- FEMA is central to recovery information coordination and owns the largest number of recovery tools in the MoDI

Each of the 320 tools in the MoDI is tagged with the phase or phases of emergency management for which it is used at the Federal level. For example, a tool may be used to support both planning and response while another tool is applicable exclusively to recovery efforts. Network and quantitative analysis were used to understand how information is exchanged between recovery tools and the datasets and models used to support other phases of emergency management, to quantify the availability of recovery tools, and to characterize the distribution of recovery tools owned by different agencies. The results are discussed below.

Network integration of recovery datasets and models

The integration of recovery tools into the overall community of datasets and models was tested by analyzing their overlap and interconnections with datasets and models used during other phases of emergency management. Response and recovery overlap in the emergency response timeline (Figure 1) and rely on common fundamental data about the event: what happened, who and what is impacted, and to what degree. As a result, it is expected that those working in the recovery mission would utilize most of the response-relevant tools in addition to recovery-specific datasets and models.

The hurricane network was used as the basis for a comparison between those tools in recovery versus those used for response. The hurricane network was selected because it is a relatively well-tested hazard with a robust and mature information network.

Contrary to expectations, most tools in the hurricane network are selectively used to support either response or recovery, but not both. As shown in Figure 2, relatively few tools are shared between response and recovery phases.[3] In addition, response tools are well integrated into the hurricane network, as shown by their relatively uniform distribution across the network map. By contrast, the majority of the recovery-specific tools are clustered together in the lower right of the hurricane network and are primarily connected to each other with relatively few connections to the more interconnected portions of the network containing the recovery-specific tools.

[3] The hurricane network contains 189 tools. 91 are 'response-specific' (tagged response and not recovery) and 39 are 'recovery-specific' (tagged recovery and not response). Tools colored gray include mostly tools that are used to support both response and recovery (44 of 59), but also tools that are not used for either response or recovery phases (15 of 59).

A number of response-specific and recovery-specific tools are completely disconnected from the hurricane network. These "orphan" tools, grouped together in the top right of Figure 2, do not share information with other datasets and models in the network. As a proportion of their total number, a much greater percentage of the recovery-specific tools are orphans than for response.

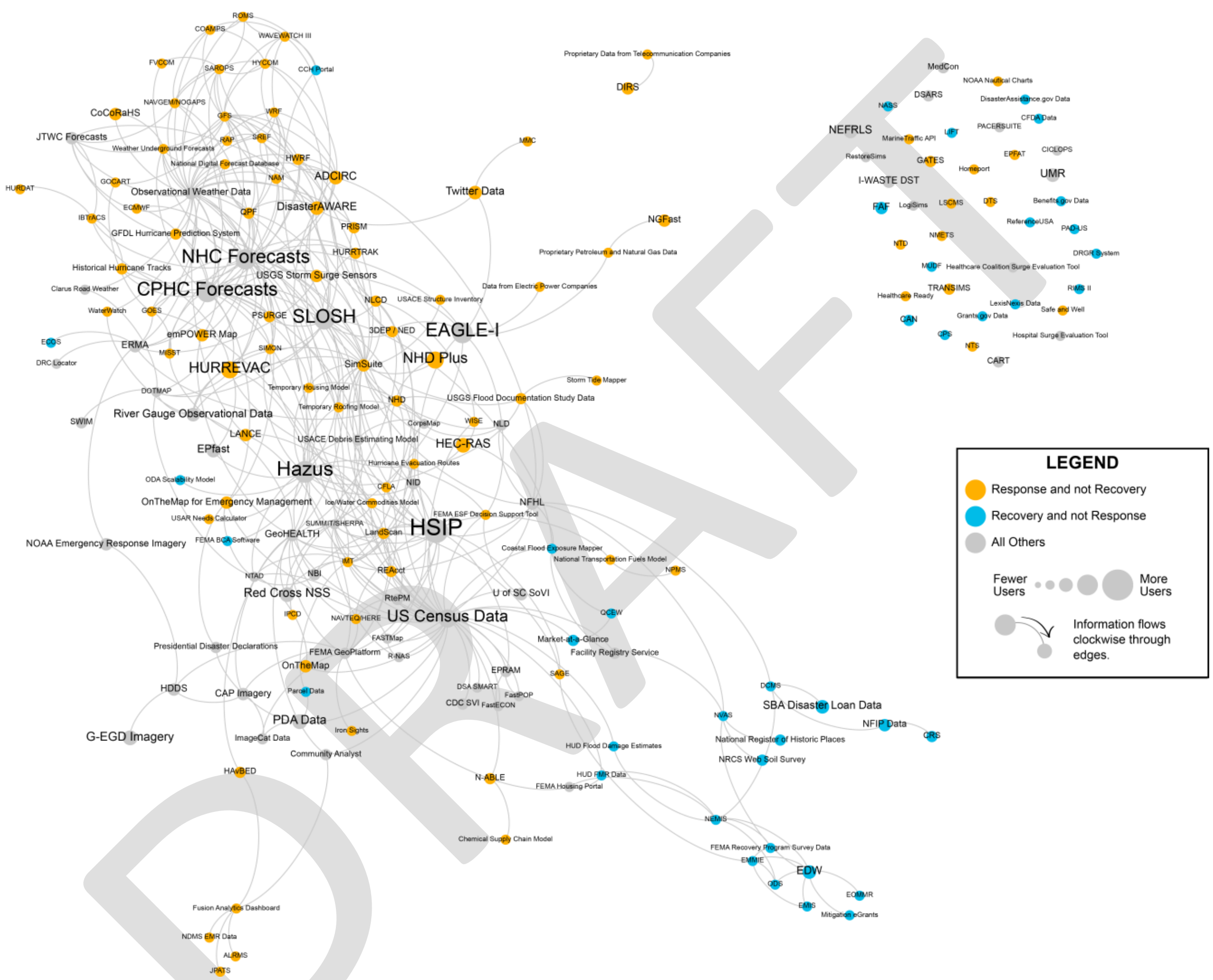

Figure 2. Response and Recovery-specific Tools Within the Network Map of Hurricane Models and Datasets. Nodes (circles) are sized by the number of federal agencies using the tool. Information flows clockwise along edges (lines) between the datasets and models and indicate data transfer between information resources. Node colored gold are used for response and not recovery while nodes colored blue are used for recovery and not response.

Availability of recovery tools

As shown above, significantly fewer tools are used to support recovery operations (95 of 320 or 30%) than response operations (293 of 320 or 92%). Indeed, as shown in Table 1, below, response tools significantly outnumber recovery tools for every hazard studied to date with the most striking disparity in biological hazard scenarios. Only 13 tools are available to support recovery for recovery following an intentional biological attack and just 8 tools support recovery from a naturally-occurring biological outbreak.

Table 1. Fraction of Tools Used to Support Response and Recovery, by Hazard

Hazard	Response-relevant	Recovery-relevant
Hurricane	71% (134 / 189)	44% (83 / 134)
Earthquake	67% (101 / 151)	48% (72 / 101)
Improvised Nuclear Device (IND)	70% (119 / 169	40% (68 / 119)
Biological-Intentional Attack	86% (64 / 74)	18% (13 / 64)
Biological-Naturally Occurring Outbreak	86% (63 of 73)	11% (8 of 63)

Agency ownership of recovery datasets and models

FEMA owns the greatest number of recovery tools in the inventory (Figure 3). These datasets and models span the flow of information from raw data to mission-specific requirements, suggesting that FEMA is the primary source for recovery information, as is consistent with the prominent role for FEMA in supporting and coordinating interagency recovery operations. Most of the other agency owners of recovery tools tend to own models and datasets toward one end of the flow of information: the National Oceanic and Atmospheric Administration (NOAA), U.S. Geological Survey (USGS), and Department of Transportation (DOT) own tools involved in event characterization and providing situational awareness while agencies such as the Environmental Protection Agency (EPA) and Small Business Administration (SBA) own tools more focused on operational requirements (decision support tools and mission specific requirement data).

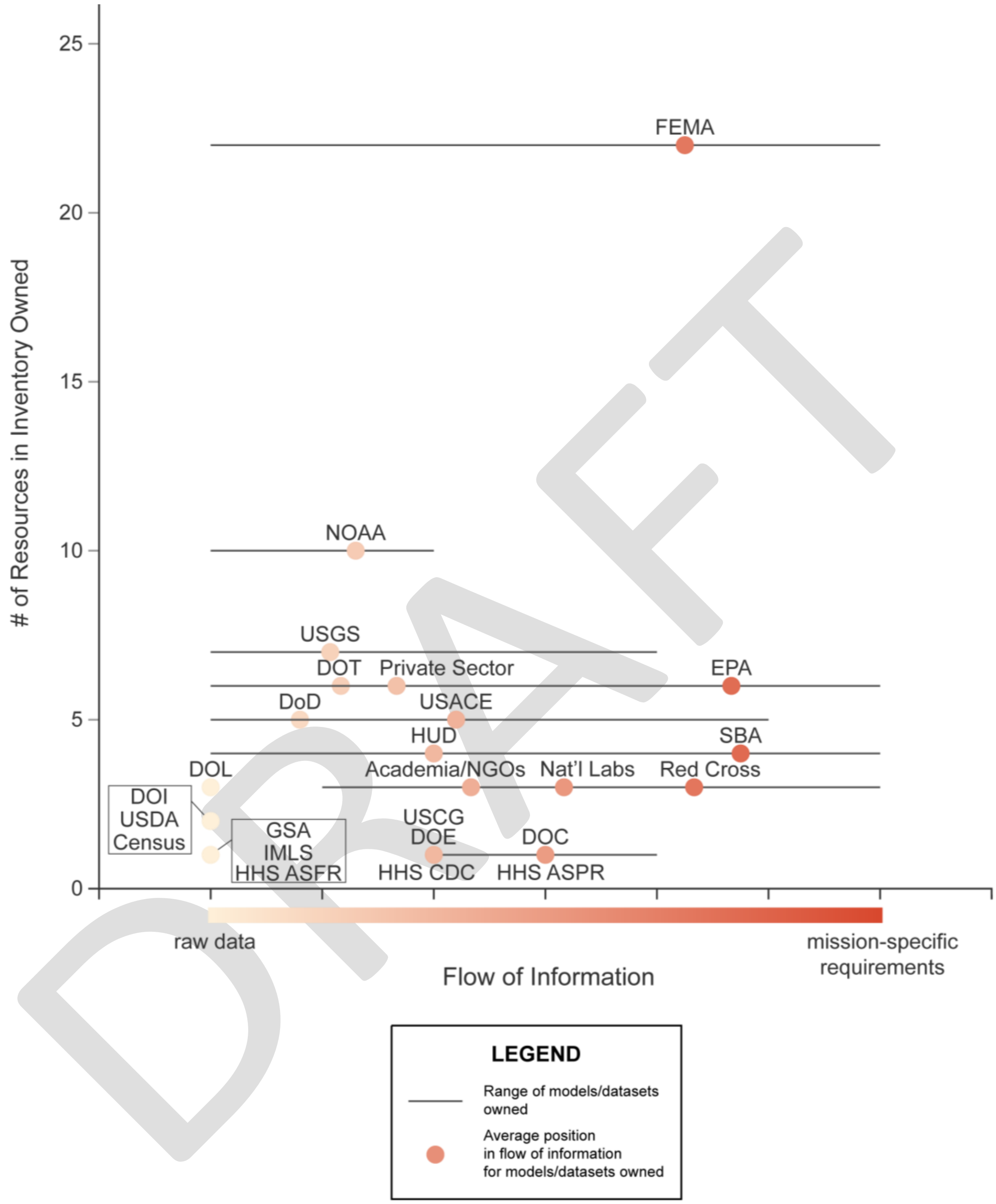

Figure 3. Dataset and Model Ownership for Recovery. Agencies are plotted with respect to the number of models and datasets they own (y-axis). The color and position of each agency on the x-axis represents an average of the types of tools that agency owns. Agencies positioned more to the right or left tend to own tools toward those respective ends of the flow of information. The horizontal bars indicate the range of tool types each agency owns.

To further analyze the distribution of recovery information types owned by each agency, a Sankey diagram shows information types on the left (organized according to flow of information framework) connected to agencies on the right, based on the information types they own (Figure 4). Notably, a broad range of agencies own recovery tools, and many agencies contribute one or a few tools to support recovery efforts, especially raw datasets and situational awareness data. However, the relatively smaller numbers of event characterization and consequence models suggests that these data are likely to remain unprocessed due to a lack of tools to support further processing into impact estimates. For example, raw datasets, such as the Protected Areas Database of the United States (PAD-US) and the Museum Universe Data File (MUDF), are critically important to the environmental and historic preservation mission, which requires identifying sites and properties impacted by the event. These raw datasets contain the locations and names of significant sites, but analysis to determine which sites were impacted and to what degree is currently performed manually and does not consistently incorporate the latest event-specific consequence analysis to determine impacts to those sites.

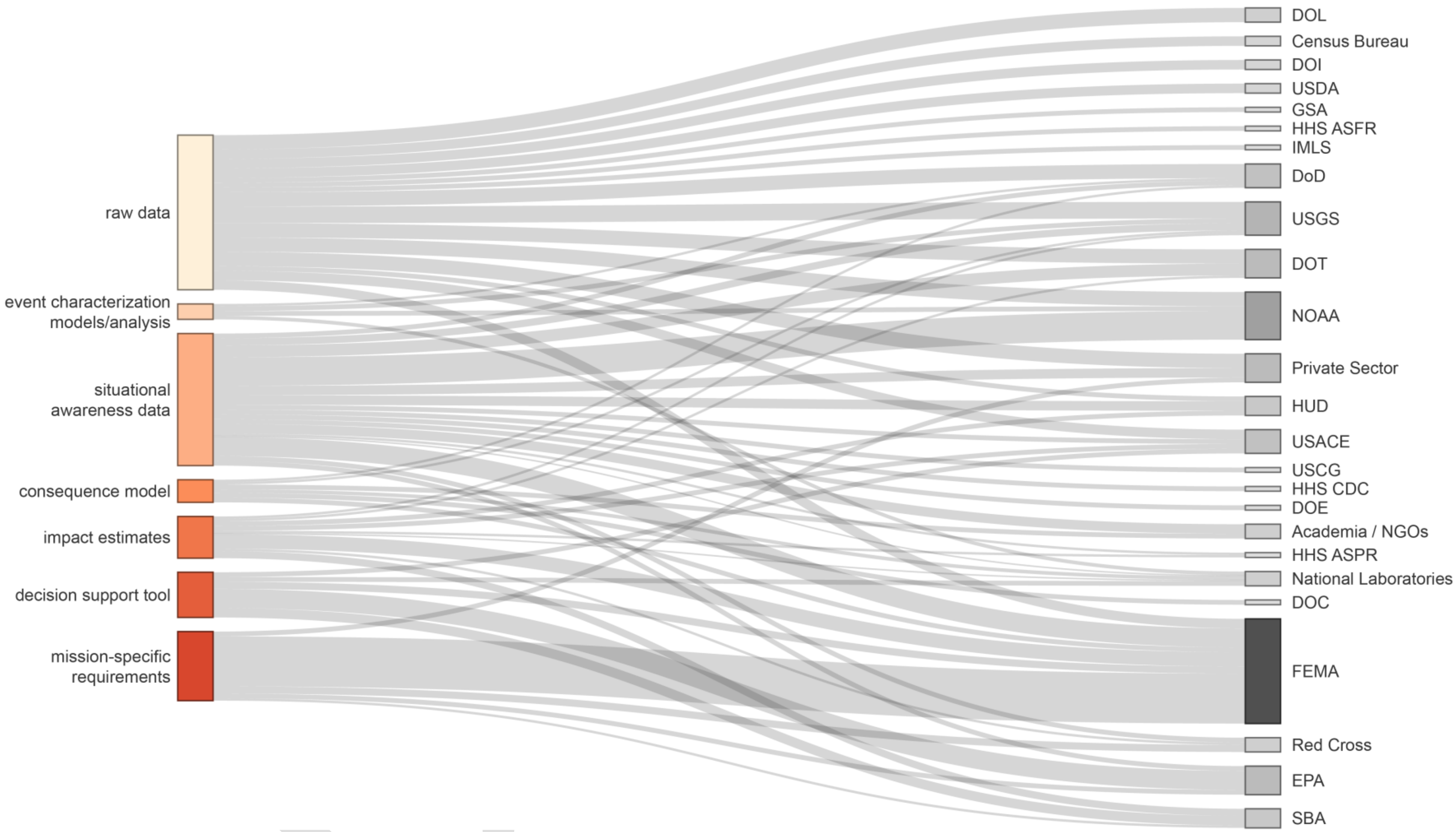

Figure 4. Recovery Owner Sankey Diagram. Agency model/dataset owners (right) are connected in proportion to their tool ownerhsip to each type of information (left). Dataset and model types are oragnized by the flow of information framework down the left side and sizes are proportional to total number of tools of each type. Agencies are sized and shaded according to the number of tools they own. For tools tagged as multiple types, multiple connecitons are included: for example, a tool tagged as raw data and situational awareness data would connect its agency owner to each of those information types.

Agency owner centrality

In addition to owning the largest number of tools for recovery, FEMA owns the set of tools most central to the recovery information network. Centrality measures how often a particular model or dataset serves as a bridge to connect other models and datasets in the network. In this way, central tools serve a key function to integrate and process information in the network. As a result, agencies owning more central tools are expected to act as hubs for recovery data processing and information coordination. FEMA has the highest average centrality (darker blues represent greater centrality) and largest number (agency size is proportional to the number of tools) of datasets and models in the owner network, consistent with its role in coordination of interagency recovery efforts. Importantly, the recovery agency owner network map specifically reflects agency recovery and roles which can differ from other phases of emergency management. For example, SBA owns highly central tools for recovery, reflecting the role for SBA in supporting recovery efforts, even though this agency does not own tools central to response.

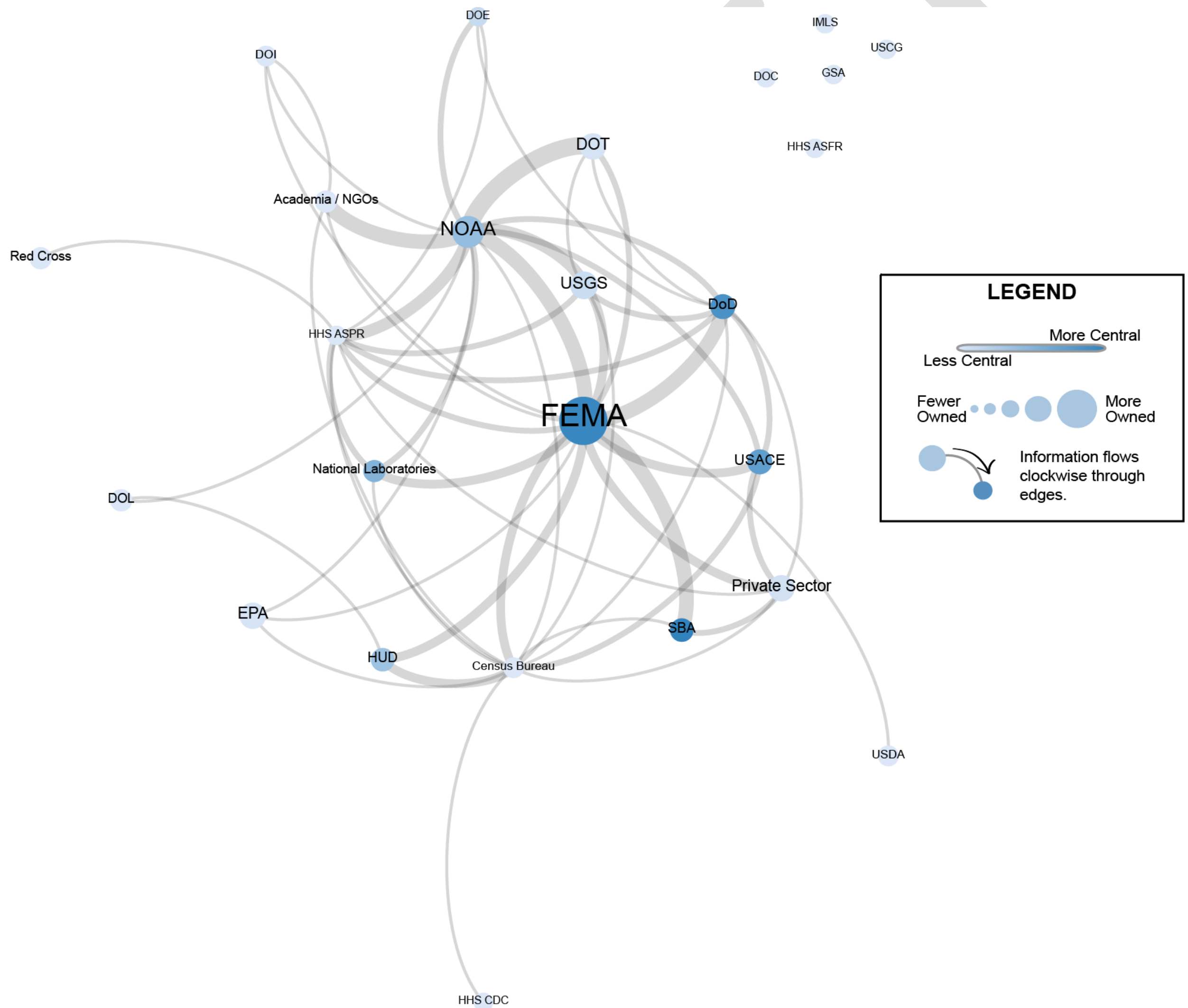

Figure 5. Centrality of Agency Model/Dataset Ownership. Nodes (circles) represent federal agencies that own models/datasets used to support recovery and are sized by the number of models/datasets owned. Information flows

clockwise along edges (lines) between the datasets and models and indicate data transfer between information resources. Each node is colored from less central (lighter blues) to more central (darker blues) based on betweenness centrality. Centrality and information flow in this network represent the average across all models and datasets owned by a given agency for recovery.

The results of this report, commissioned by the MDWG, provide the first analysis of the datasets and models used to support a specific phase of emergency management, rather than taking a hazard-centered approach. Quantitative analysis demonstrated that fewer models and datasets are used to support recovery operations compared to response across all hazards included in the MoDI. Quantification of the recovery tools used, in combination with information garnered from interviews, suggested that many raw data are not further processed into formats that directly support recovery data requirements. Using the hurricane network as a prototype example, recovery models were shown to be less integrated into the overall information network, as compared to recovery tools. Though FEMA is a central owner of tools for recovery, the agencies that contribute to interagency recovery response are numerous and diverse.

Summary and Recommendations

Here, the datasets and models specifically used by the recovery mission within the federal emergency management community were identified and analyzed to better understand how information is exchanged between those tools and those datasets and models used for other phases. Information gathered from interviews and subsequent analysis revealed two key findings: a lack of integration of recovery information with response tools, including a shortage of recovery-specific models to process raw datasets, and marked lack of information sharing between event phases and agencies.

Improve integration of response and recovery

Recovery tools are not well-integrated with inventory tools used for other phases, as shown in the hurricane network map (see Figure 5). Therefore, recovery tools are largely disconnected from the upstream event characterization data developed to support the response phase, including the real-time event characterization that provides the most up-to-date information about the event. In some cases, better publicizing and centralizing the distribution of response datasets and model outputs can make them readily accessible and useable by the recovery community. In other cases, the inputs of recovery tools, or the outputs of response tools, should be modified so that the tools integrate common information.

Development of new models or adapting response models for use during recovery represents an opportunity to improve translation of existing data into information that supports mission-specific requirements for recovery. For example, many recovery tools do not automatically integrate imagery or damage assessments post-event and, instead, rely on manual data processing and ingestion. For example, no automated connection between preliminary damage assessment (PDA) data, collected to assess on-the-ground impacts immediately post-event, and recovery tools currently exists. PDA data are initially collected during response, but have been described as inconsistently available for subsequent use in the recovery phase, or as needing to be recollected. Furthermore, a lack of event characterization and consequence models to support recovery contributes to a large number of unprocessed raw datasets and situational awareness data. Many agencies own useful unprocessed datasets that would improve recovery information if integrated into the network of other tools. For example, several datasets that support the environmental and historic preservation mission are not currently connected to the recovery network, but provide information essential to recovery operations. Targeted investments in model development or dataset and model integration represent an opportunity to fill these gaps in information flow.

Develop improved approaches to information sharing

Barriers to integration of recovery information can be improved by connecting currently unconnected datasets and models, but can also be improved through enhanced data sharing and exchange within the interagency. Response and recovery communities should focus on the information handoff during events from response-focused operations in the Joint Field Office (JFO) to the recovery efforts led by the Federal Disaster Recovery Coordinator (FDRC) and from ESFs to RSFs. For example, both response and recovery operations rely on information about the current status of transportation systems and debris removal and ensuring the handoff of the best available response information has the potential to best support early recovery efforts. Subsequently, the agencies with a primary role in initial recovery, such as the Individual Assistance program at FEMA, can inform longer-term recovery. However, as the individual assistance example highlights, recovery data contain sensitive personally identifiable information (PII). Agencies make a priority of finalizing the information sharing agreements that protect this information, for example by anonymizing and rolling up information to a non-PII level, while sharing the best available information across recovery missions.

Agencies often described their tools in relation to the highly central and widely used recovery tools owned by the central information coordinator for recovery, FEMA. Though some data is successfully coordinated through FEMA, data sharing between agencies remains a challenge. Technical barriers exist to making enterprise-level data centers “talk to each other”: there are few standards for data formats, and interviewees identified data sharing agreements that have languished in the development stages, even for decades. For example, the Department of Housing and Urban Development (HUD) Disaster Recovery Grant Reporting (DRGR) System, and similar systems in use for grant management at FEMA, the SBA, and the Department of Labor (DOL), are not currently linked through an API or other automated data sharing system. More effective information sharing between these systems would provide the means for recovery operations personnel to consistently access data needed to support interagency recovery activities.

Conclusion

Analysis of the datasets and models used to support Federal recovery operations, as performed here, highlights the diversity of recovery tools and the agencies that own them. Network analysis, combined with information from interviewees, suggests that recovery tools are not well integrated with datasets and models used to support other phases of emergency management, particularly recovery. Improving connections among recovery tools, and between recovery tools and those used for other phases, represents an opportunity to fill gaps in the information currently available to support recovery. In addition, recovery information availability could be improved by focusing on the handoff between response and recovery information and in removing barriers to information sharing across agencies.

Appendix A: Interviewees

Name	Agency or Organization
McNamee, Shannon	American Red Cross
Betz, Dawn	Digital Globe
Kralovec, Matthew	Digital Globe
Asadurian, Alis	Department of Commerce Economics and Statistics Administration (DOC ESA)
Cooke-Hull, Sandra	DOC ESA
Henry, David	DOC ESA
Hait, Andy	DOC ESA Census Bureau
Walker, Chip	DOC ESA Census Bureau
Zamora-Appel, Barbara	DOC ESA Census Bureau
Mitchell, Cedric	Department of Defense (DoD)
Dykas, Walter	Department of Energy Office of Electricity Delivery and Energy Reliability (DOE OE)
Greenberg, Jeremy	Department of Transportation (DOT)
Ridge, Matt	DOT
Murray, Alexander	DOT Federal Aviation Administration (DOT FAA)
Strocko, Ed	DOT Office of the Assistant Secretary for Research and Technology (DOT OST-R)
Canzler, Erica	Environmental Protection Agency Office of Emergency Management Consequence Management Advisory Division (EPA OEM CMAD)
Snyder, Emily	EPA Office of Research and Development National Homeland Security Research Center (EPA ORD NHSRC)
Battle, Ashley	Federal Emergency Management Agency (FEMA)
Carrol, Rebecca	FEMA

Janca, Andrew	FEMA
Koziol, Lauralee	FEMA
Potosnak, Ryan	FEMA
Tinsman, Mark	FEMA
George, Kysa	FEMA Office of External Affairs (FEMA EA)
Robbins, Ronald	FEMA EA
Springer, Jody	FEMA Federal Insurance and Mitigation Administration (FEMA FIMA)
Moline, Julia	FEMA Field Operations Directorate (FEMA FOD)
Bennett, Gerilee	FEMA FOD Interagency Coordination Division (FEMA FOD ICD)
Tomic, Leslie	FEMA FOD ICD
Paulette-Chapman, Dan	FEMA National Preparedness Assessment Division (FEMA NPAD)
Thomas, Claire	FEMA NPAD
Foley, Lori	FEMA Office of Environmental Planning and Historic Preservation (FEMA OEHP)
Pettit, Kimberly	FEMA OEHP
Brierley, Mick	FEMA Office of Response and Recovery (FEMA ORR)
Campbell, Matt	FEMA ORR
Criswell, Deanne	FEMA ORR
Edgar, Gregory	FEMA ORR
Hale, Clay	FEMA ORR
Jawaid, Naveen	FEMA ORR
Krzystyniak-Sobiewska, Anna	FEMA ORR
Meyer, Emily	FEMA ORR
Souza, Kevin	FEMA ORR
Warren, David	FEMA ORR

Zuzak, Casey	FEMA ORR
Bonifas, Michelle	FEMA ORR Recovery Analytics Division (FEMA ORR RAD)
Faber, Chad	FEMA ORR RAD
Jediny, John	General Services Administration (GSA)
Hopper, Ken	Department of Health and Human Services Office of the Assistant Secretary for Preparedness and Response (HHS ASPR)
Pereira, Esmerelda	HHS ASPR
Ziaya, David R.	Department of Housing and Urban Development Office of Administration (HUD OA)
Richardson, Todd	HUD Office of Policy Development and Research (HUD PDR)
Eslinger, Sandy	National Oceanic and Atmospheric Administration Office for Coastal Management (NOAA OCM)
Bausch, Doug	Pacific Disaster Center (PDC)
Green, Joe	PDC
Dial, Patrick	Small Business Administration (SBA)
Brumefield, Tyrone	United States Army Corps of Engineers (USACE)
Sheeley, Jason	USACE
Simrall, Robert	USACE
Boyd, Valerie	United States Coast Guard (USCG)
Nguyen, Jason	USCG
Barnes, Joshua	White House National Security Council (NSC)

Appendix B: Methods

The workflow of data collection, processing, and analysis performed for this project is shown in Figure A1 and described briefly in the Methods section in the main text. Each step is described in detail below.

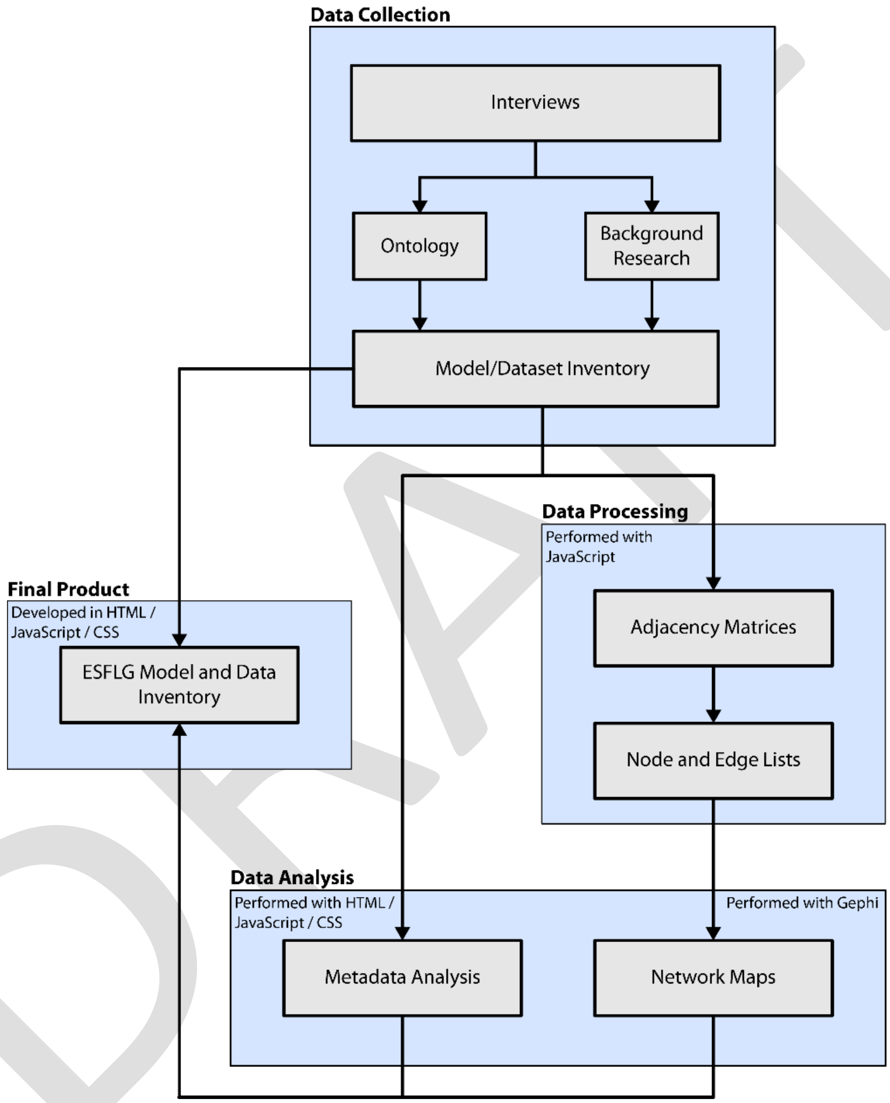

Figure B1. Analysis workflow. A depiction of the sequence of work involved in producing quantitative analysis of the model/dataset inventory.

Data Collection

Interviews

The information required to analyze the available data and modeling tools was collected through a series of in-person and phone interviews. Interviews were performed with the members of the MDWG, the subject matter experts and stakeholders they recommended, and additional individuals suggested by those interviewed. During these interviews, the users and owners of each tool identified and characterized the ways in which it is used to support planning and operational decision making. In most cases, the MDWG members were interviewed initially. Interviews with additional subject matter experts or leadership were scheduled upon recommendation to provide further breadth or depth of information, depending on the size of the agency or division represented and the expertise of each interviewee.

Interviews were opened with an introduction to the project. Throughout the project, interviewees have included those who are providers of data or are tool developers; those who are analysts and users of those data and tools; those who make operational decisions informed by datasets and models; and those who have roles that include a combination of tool development, analysis, and decision making. Interviews were designed to capture an overview of the roles and responsibilities of each group and the ways in which data and data processing tools, including modeling, supported those roles. The flow of the conversation varied widely based on the expertise of the interviewee and attempted to capture both the general and specific information requirements from each interviewee across the spectrum of emergency management missions and the phases of an emergency. A comprehensive list of the interviewees can be found in Appendix A.

Ontology

Initial iterations of the MDWG effort were focused specifically on hurricane and earthquake hazards. Interviewees were asked about the types of information they need to support their emergency management mission in the context of those hazard scenarios. Based on the responses, a systems-level analysis of the information requirements was conducted and a framework to capture the flow of information between the different types of data and modeling tools was developed. This ontology describes how the information required is collected and processed over several iterations of collation and analysis. This analysis provides a framework to understand the role and value of both computationally intensive predictive modeling and the rapid calculations provided by simple algorithms to determine mission specific requirements. This information ontology or flow of information framework was vetted and validated by the working group and is described in detail in the main text of this report. It has been used to describe the flow of information in support of emergency management for all hazards analyzed since the initial hurricane and earthquake efforts, which now also include flood, nuclear detonation, and biological hazard scenarios.

Model/Dataset Inventory

A comprehensive inventory of tools used across the federal interagency and the linkages between them was generated on the basis of the tools discussed during interviews, followed by background research to identify inputs and outputs of each tool. Only tools with federal users were included in the inventory. Tools under development or not currently used to support emergency management activities were identified, but not included in the inventory. Information about these tools and how they function within the flow of information has been retained in an archived library. This information allows for more a more detailed analysis and verification of the analyses. Additionally, these tools can be used in future to suggest mechanisms to fill gaps identified in the current inventory. The inclusion of only used and operational tools in the inventory enables an analysis of how information currently travels within the interagency and results in a streamlined inventory containing the information immediately useful for emergency managers.

Metadata

The flow of information framework captures the functional, time-dependent, and mission-specific variation between tools used across the federal interagency. However, it does not describe other essential characteristics, such as how those tools are accessed, used, and updated. These additional characteristics, or metadata, must also be collected to properly organize and analyze the tools to maximize effective usage during all phases of emergency management. These metadata will appear in the interactive inventory of tools, the Model and Data Inventory (MoDI), upon completion of the project.

Metadata categories include: the tool's name; summary; resource type; applicable hazards; supported Core Capabilities, Emergency Support Functions, and Recovery Support Functions; keywords; data collection method; owner; users (federal agency-level) by hazard and phase; upstream inventory resources by hazard; downstream inventory resources by hazard; phase specific utility by hazard; access information; processing requirements; refresh rate; last known version; programming language; output file types; technical contact; real-time contact; geographic coverage; and website. Complete descriptions of each metadata category are included in the MoDI.

As metadata were collected and input into the inventory, scripts written in the R language were used to automatically populate certain metadata categories based on the contents of other metadata categories. For example, certain Core Capabilities were automatically tagged based on the Emergency Support Functions and Recovery Support Functions that were tagged for the tool. R was chosen because of its flexibility and suitability for repetitive text processing tasks. [4]

In addition to the data collection methods shown in Figure B1, a data validation step was performed. After the initial drafts of analyses and the inventory were completed, the metadata gathered from interviewees describing tools they own or use were submitted to those interviewees for review and feedback. All feedback received was adjudicated and incorporated into the inventory and results.

Data Processing

In the data processing phase, the model/dataset inventory data were processed into a format that could be imported and analyzed by network analysis software. All data processing was performed using JavaScript. JavaScript was chosen because the inventory data could be readily represented in JavaScript Object Notation (JSON) format, which is widely used in data visualization and web-based applications. Additionally, many robust open source libraries are available for JavaScript-based data analysis, facilitating reuse of the analysis scripts and their transfer to other parties.

Networks

A network is defined as a system consisting of interconnected components. Network analysis is the process of understanding the connections between those components. Individual components of the

---

[4] R Development Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org

network are called nodes, and the connections between them are called edges, with information moving through the network by a defined, or directed, flow. Networks can be represented by objects called adjacency matrices, node lists, and edge lists, as described in the following sections.

Adjacency Matrices

To build network maps describing the linkages between tools in the inventory, the metadata defining the upstream and downstream linkages between tools were used to build an adjacency matrix. An adjacency matrix is a mathematical method of representing a network that provides a simple way to calculate many network measures and statistics.[5]

Node and Edge Lists

The adjacency matrix was then converted into an edge list. An edge is a line in the network that connects two nodes, and in this case, represents the transfer of information from one tool to another. The edge list contains a list of connections between nodes in the network. In addition, the inventory metadata were used to prepare a node list. A node is a point in a network, and in this case, each node represents a single tool in the inventory. The node list contains the metadata of each node in the network, allowing that information to be visualized on a network map and analyzed in the context of the network. These node and edge lists were imported into Gephi,[6] an open source network visualization and analysis software program, to create and analyze the network maps used in the analysis.

Data Analysis

The inventory data, including the tools and their associated metadata, and the networks based on this inventory, were used to perform an analysis of the inventory, as described in the results section. Two main types of analyses were performed: network analysis and metadata analysis.

Network Analysis

The majority of network analysis presented in this report describes the connections between the datasets and models used by the federal interagency in the context of emergency management. Two metadata categories (upstream inventory resources and downstream inventory resources) describe linkages between the tools based on the flow of information between those tools. These linkages were used to build flow-based tool networks. Additionally, each tool was tagged with a federal agency owner. This ownership information was used to build agency networks showing the flow of information between federal agencies, based on the tools owned by each agency and the linkages between those tools.

Tool Networks

Tool network maps were generated in order to visualize and analyze the connections between tools described in the inventory. A simple, notional example of a tool network map is shown in Figure B2. In tool network maps, each node represents a single tool in the inventory. The size of a tool's node and its label is proportional to the number of federal agency-level users of the tool. Here, the number of users of a tool is defined as the total number of federal agencies that directly use the tool in the context of

---

5 A short, rigorous definition of an adjacency matrix: For a network of n nodes, the adjacency matrix A is an n x n matrix where the i,j^{th} entry in the matrix represents the number of connections from the i^{th} node in the network, to the j^{th} node in the network.

6 Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

their work supporting interagency emergency management efforts. Edges connecting nodes in the tool network represent the flow of information and processing of data as it passes from one tool to another. Information flows in a clockwise direction, with edges curving clockwise from tools that act as source of information toward tools that consume that information. Both the inputs (upstream tools) and outputs (downstream tools) of each tool were identified based on in-depth analysis of interview data and a review of the technical documentation of the tool, when available.

Nodes were arranged by a force direction algorithm that groups closely linked nodes.[7] This algorithm treats each node as a charged particle that repels all other nodes, and each edge as a spring, pulling the connected nodes back together. The ForceAtlas2 implementation of the force direction algorithm was used, as made available in Gephi. The force direction algorithm was chosen because it has been shown to highlight underlying community structures that exist in data.[8] The force direction algorithm was run and re-run numerous times, with the initial positions of nodes randomized at the beginning of each run. Additionally, the strength of repulsion between nodes was varied from weak to strong. Randomly restarting the algorithm and systematically varying node charge allowed the structure of the graph to be fully explored under a variety of conditions. The tool network maps shown in the results section are the result of multiple runs of the force direction algorithm that consistently converged on a particular layout of the nodes.

Each tool network map used a specific node color scheme to convey information. Tool network maps used one of three color schemes: resource type, betweenness centrality, and modularity cluster, as described in the following sections.

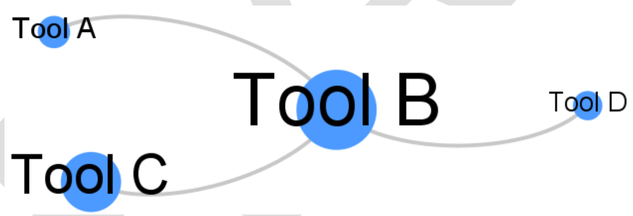

Figure B2. Example of a simple tool network map. Individual tools are represented by blue discs (nodes). Direct connections between tools are represented by gray curved lines (edges). The flow of information travels clockwise. In this example, information flows into Tool B from Tools A and D. Information from Tool B flows into Tool C. The size of each node can convey additional information. For the tool network maps presented in this report, node sizes are proportional to the number of users of that tool.

Resource Type

The connections between tools in the context of the flow of information ontology were examined by creating tool network maps with nodes colored according to their position in the flow of information, or resource type. There are seven possible resource types, as described in the main text. Lighter-colored nodes represent tools that are upstream in the flow of information (relatively unprocessed information), and darker-colored nodes represent tools downstream in the flow (relatively processed information).

7 Jacomy, M., Venturini, T., Heymann, S., Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE*, *9*(6), e98679.

8 Noack, A. (2009). Modularity clustering is force-directed layout. *Physical Review E*, *79*(2), 26102.

Tools that function as more than one resource type are colored based on the average position of those types in the flow of information.

Betweenness Centrality

Centrality is a family of measures used to rank the most significant nodes in a network. Different definitions of centrality have been proposed to measure particular qualities of a node's position within a network, and the most prevalent definitions can be classified into one of three groups: degree, closeness, and betweenness.[9] Degree centrality is based on the number of connections that one node has to other nodes, whereas closeness centrality is a function of a node's distance from other points in the network. Here, the significance of individual tools was investigated with a measure of betweenness centrality, which ranks nodes according to the frequency with which they lie between other nodes in the network. Betweenness centrality was analyzed because it provided the means to determine which tools act as information "bridges" between other tools. This "bridge"-like character of a node cannot be readily evaluated by visual inspection of a force-directed network map, whereas properties such as degree and closeness may be.

The relative importance of specific tools as information "bridges" was investigated using the betweenness centrality measure. Betweenness centrality is a common metric of node significance that characterizes how often a node lies between other nodes in a network. Here, the betweenness centrality of a node is defined as the sum of minimum-length paths between other nodes that an individual node lies on, with each path weighted according to the inverse of the number of alternative same-length paths between the corresponding node pair.[10] Only paths in the direction of information flow are considered. High betweenness centrality is assigned to nodes that act as "shortcuts" or "bridges" between different parts of the network, and is an estimate of the amount of information flowing through a node relative to other nodes. [11,12] In the network maps, nodes were colored on a gradient such that more central nodes were darker and less central nodes were lighter.

Betweenness centrality only considers the shortest paths between nodes and therefore does not consider longer, alternative paths over which information could be passed within a network. Here, the weighted version of the betweenness centrality calculation was used in order to highlight the significance of nodes that act as the only information "bridge" between other nodes. These nodes represent tools with high information "bridge" character that, if defunded or removed from operational use, could lead to a breakdown in the flow of information between tools.

Additionally, the betweenness centrality of each node may be represented either as its absolute betweenness centrality score, or as its score normalized relative to the score of the most central node. The latter option was chosen so that the most and least central tools in each network map would be apparent from node coloring regardless of absolute betweenness centrality.

---

9 Newman, M. (2010). *Networks: an introduction*. Oxford University Press.

10 Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, *25*(2), 163–177.

11 Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry*: 35-41

12 Freeman LC (1979) Centrality in Social Networks Conceptual Clarification. *Social Networks* 1: 215-239

Agency Networks

In addition to tool network maps, agency network maps were created to analyze the flow of information between tools owned by different agencies. A simple example of an agency network map is shown in Figure B3. In agency network maps, each node represents a single federal agency-level owner in the inventory. The size of an agency's node and its label is proportional to the number of tools owned by the agency. Edges connecting nodes in the agency network represent the flow of information from a tool owned by one agency into a tool owned by a different agency. As with tool networks, information flows in a clockwise direction. The width of each edge is proportional to the number of tools owned by the source agency that feed information into a tool owned by the target agency. Self-edges, or circular edges from a node into itself, are not shown.

As described previously for tool networks, nodes were arranged by a force direction algorithm that groups closely linked nodes. Each node in the agency network map was colored based on the average betweenness centrality score of the tools it owned in the corresponding tool network map.

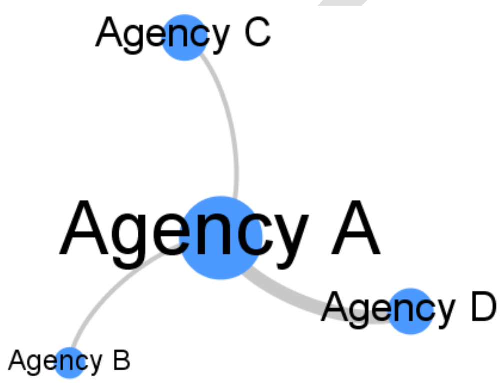

Figure B3. Example of a simple agency network map. Individual agencies are represented by blue discs (nodes). Gray curved lines (edges) represent connections between tools owned by different agencies. The flow of information travels clockwise. Edge thickness is proportional to the number of tools owned by the source agency that feed a tool owned by the target agency. In this example, Agency B and Agency C each own one tool that feeds a tool owned by Agency A. Agency D owns three tools that feed a tool owned by Agency A, so the edge connecting these two agency nodes is thicker. The size of each node can convey additional information. For the agency network maps presented in this report, node sizes are proportional to the number of tools owned by that agency.

Metadata Analysis

In addition to network analysis of the tools' linkages, quantitative analyses of inventory metadata were performed to characterize the tools available to support emergency management in terms of other attributes. These analyses included the number of tools owned by each agency, by resource type; the absolute number and percentage of orphan tools (tools with no connections to other tools), by hazard; the absolute number and percentage of tools available for each phase; and the most-used tools, by hazard. These analyses were presented in tabular form, or visualized in bar graphs prepared with the d3.js framework.[13]

Additional, more sophisticated visualizations were prepared for some metadata analyses, including owner dot-and-dot-and-whisker plots and Sankey diagrams, described in the next sections.

Owner Dot-and-whisker plots

Owner dot-and-whisker plots were created to visualize the number of tools owned by each agency, the average resource type of those tools, and the range of resource types spanned by those tools.

In the plots, each owner agency is represented by a dot. Each dot's vertical position is determined by the number of tools owned by the federal agency, with agencies owning more tools being positioned higher on the vertical axis. Each dot's horizontal position is determined by the average resource type of the tools owned by that agency, with agencies owning tools that are on average more upstream in the flow of information positioned toward the left, and downstream positioned toward the right. The color of each dot represents the average resource type of the tools owned by that agency, with lighter dots representing agencies owning tools that are on average more upstream in the flow of information, and darker dots more downstream. Finally, a black horizontal line transects each dot, with its left and right end points defined by the most upstream and most downstream resource type owned by that agency. Agencies owning only tools with only one resource type do not have horizontal lines.

Owner dot-and-whisker plots were created using custom JavaScript scripts based on the d3.js framework.[14]

Sankey Diagrams

Sankey diagrams were used to visualize the number of tools of each resource type owned by different federal agencies. Sankey diagrams are flow diagrams that show what components make up the whole of a particular element, and in what proportions.

Here, the Sankey diagram consists of two sets of rectangles. The left set of rectangles each represent one of the seven resource types, with the height of each rectangle proportional to the number of tools tagged with the resource type. Each resource type rectangle is colored based on its position in the flow of information: resource types upstream in the flow are lighter, and types downstream in the flow darker.

The right set of rectangles represent the federal agencies that own tools, with the height of each rectangle proportional to the number of tools owned by the agency. Each agency owner rectangle is colored based on the number of tools it owns: agencies owning fewer tools are lighter, and agencies owning more tools are darker.

13 Bostock, M. (2012). Data-Driven Documents (d3.js), a visualization framework for internet browsers running JavaScript. Retrieved from http://d3js.org/

14 Bostock, M. (2012). Data-Driven Documents (d3.js), a visualization framework for internet browsers running JavaScript. Retrieved from http://d3js.org/

The bands signify what proportion of the tools owned by each agency are tagged with each resource type. The thickness of the band between an agency rectangle and a resource type rectangle is proportional to the number of tools that agency owns that are tagged with that resource type.

The library used to create Sankey diagrams is a plug-in for the d3.js framework called Sankey.js.[15]

[15] Bostock, M. (2012). Sankey Diagrams. Retrieved from https://bost.ocks.org/mike/sankey/