SANDIA REPORT
SAND2018-8362
Unlimited Release
Printed July 2018

SNL Lesson Learned and Guidance for Data Repositories and Analytic Frameworks

Alisa Bandlow, Katherine A. Jones, Vanessa N. Vargas

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology and Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

SAND2018-8362
Printed July 2018
Unlimited Release

SNL Lessons Learned and Guidance for Data Repositories and Analytic Frameworks

Alisa Bandlow, Katherine A. Jones
Operations Research and Computational Analysis

Vanessa N. Vargas
Resilience and Regulatory Effects

Sandia National Laboratories
P. O. Box 5800
Albuquerque, New Mexico 87185-MS1188

Abstract

This document represents the results of deliverable D05.02 (Identify relevant efforts at SNL and other institutions) under the activity area Relevant Efforts Review. The goal of the Relevant Efforts Review activity is to identify relevant data integration efforts at SNL and possibly other institutions and compile lessons learned that are relevant to the development of a framework for data integration efforts in support of analysts and decision makers. The intent of this activity is to provide, by examples, context of how the requirements-gathering process has already been implemented in other instances and to guide the development of such a process for OCIA's needs. Information for this report was gathered through SNL staff interviews and the team members' knowledge and project experiences.

TABLE OF CONTENTS

Contents

1. Scope....9
2. Background and Motivation....9
3. Data management....11
   - 3.1. Funding....11
   - 3.2. Data Management Roles....11
     - 3.2.1. Data Manager Responsibilities....11
4. Analysis and Results....15
   - 4.1. User Roles....15
   - 4.2. Methodology and Assumptions....17
     - 4.2.1. Methodology....17
     - 4.2.2. Assumptions....17
5. Examples....19
   - 5.1. Data Repository: Model and Data Inventory (MoDI)....19
   - 5.2. Data Repository: Homeland Security Enterprise (HSE) Geospatial Concept of Operations (GeoCONOPS)....20
   - 5.3. Analytic Framework: Economic Scenario Analysis....21
     - 5.3.1. Step 1: Translating scenario condition to direct impacts to critical infrastructure....21
     - 5.3.2. Step 2: Estimating infrastructure interdependencies....22
     - 5.3.3. Step 3: Estimating direct impacts from scenario and infrastructure impacts 23
     - 5.3.4. Step 4: Estimating total (direct plus indirect) impacts....23
     - 5.3.5. Step 5: Quantifying uncertainties and validating results....24
     - 5.3.6. Step 6: Reporting results....24
   - 5.4. DHS Emergency Support Functions (ESF) Decision Support Tool (DST)....25

FIGURES

Figure 1. Workflow where the advanced and view-only users use different toolsets. 16
Figure 2. Workflow where advanced and view-only users use the same analytic tool. 16
Figure 3. Speed of infrastructure dependencies for select critical infrastructure. Legend: i = instantaneous dependence, mi = minutes, h = hours, d = days, mo = months. [2] 22
Figure 4. Select critical infrastructure dependencies. [2] 23

NOMENCLATURE

Abbreviation	Definition
Abbreviation	Definition
DBA	Database administrator
DHS	Department of Homeland Security
DST	Decision Support Tool
ESF	Emergency Support Function
ESFLG	Emergency Support Function Leadership Group
FEMA	Federal Emergency Management Agency
GDP	Gross Domestic Product
GeoCONOPS	Geospatial Concept of Operations
GIOT	Geospatial Interagency Oversight Team
HIFLD	Homeland Infrastructure Foundation-Level Data
HSE	Homeland Security Enterprise
HSIP	Homeland Security Infrastructure Protection
IND	Improvised nuclear device
JIT	Just in time
MDWG	Model and Data Inventory
MoDI	Modeling and Data Working Group
NISAC	National Infrastructure Simulation and Analysis Center
OCIA	Office of Cyber and Infrastructure Analysis
POC	Point of contact
SNL	Sandia National Laboratories
URL	Uniform resource locator
U.S.	United States

1. SCOPE

This document represents the results of deliverable D05.02 (Identify relevant efforts at Sandia National Laboratories (SNL) and other institutions) under the activity area Relevant Efforts Review. The aim of the overall effort is to develop a framework to build a common operating system which addresses key questions to drive decision making and to pilot this framework through development of a prototype dashboard. The goal of the Relevant Efforts Review activity is to identify relevant data integration efforts at SNL and possibly other institutions and compile lessons learned that are relevant to the development of a framework for data integration efforts in support of analysts and decision makers. The intent of this activity is to provide, by examples, context of how the requirements-gathering process has already been implemented in other instances and to guide the development of such a process for the Office of Cyber and Infrastructure Analysis' (OCIA) needs.

Information for this report was gathered through SNL staff interviews and the team members' knowledge and project experiences.

2. BACKGROUND AND MOTIVATION

OCIA has access to a variety of datasets that can assist decision makers when planning for or responding to events impacting infrastructure. These datasets are typically not integrated to facilitate situational awareness and instead require that an individual access them separately and determines their usefulness on a more ad hoc basis. OCIA would like to better understand the key questions, tools and data which can better support decision makers. The goal is to improve the situational awareness of both OCIA analysts and the decision makers.

3. DATA MANAGEMENT

This section presents guidance on the primary roles and strategic activities that will support a successful, long-term data management system.

3.1. Funding

In addition to the funds required to purchase datasets, adequate annual funding is required to maintain a data repository. These costs include the support personnel, maintenance and upgrades for machines/servers, and software licenses and upgrades.

3.2. Data Management Roles

In an ideal world, the three roles below are separate roles fulfilled by different people. When the data repository and project scope are small, it is possible for one person to fulfill all the roles. As the data repository grows, the project scope will also grow to the point where each role becomes its own full-time job.

A **data manager** is dedicated to the role of data management and accountable for the task. The data manager is responsible for setting the strategic vision of what data is acquired, where the data is stored, and how all the data sets connect.

The **system administrator** focuses on the acquisition, maintenance, and operations of machines and servers that support the data repository and potentially the tools and models that use the data.

The **database administrator** (DBA) will focus on the acquisition, maintenance, and operations of the database environment.

3.2.1. Data Manager Responsibilities

The National Infrastructure Simulation and Analysis Center (NISAC) data manager provides the following advice on how to be successful in this role. At a minimum, this person should have prior experience as a database administrator (DBA). This person must be trusted because they will be able to see all data, some of which may be sensitive or classified. However, this person must also understand that there will be occasions when they will not be allowed to access specific datasets.

A standardized process for data management will make the entire venture run more smoothly. Proper data management is too complicated and large scale to be done in an ad hoc manner.

The data manager is responsible for developing an **acquisition plan**: what is needed, when is it needed, and how much can be spent. The data manager must have adequate time and funding to research the dataset of interest. Data research is important to understand the contract terms, to avoid procuring a redundant data set, and to avoid buying a data set when a free or cheaper option is available.

The data manager is responsible for understanding the **contract terms**.

- What is the expected scope of use for the dataset? How widespread is the data going to be used? Can the data be shared external to the purchasing agency? These answers can impact which datasets and the types of licenses procured. Some datasets may have a flat fee, while other licenses are based on the number of users ("seats"). Sometimes it can be cheaper to have named users versus generic seats.
- For one dataset, NISAC was only allowed to share data in reports with the Department of Homeland Security (DHS) if DHS also bought the dataset.
- As an extreme example, one data provider required payment for each data user, every report that contained their data, and every person who would read a report containing their data.
- There is a potential to lose existing datasets due to end of contract terms: Sometimes the end of contract stipulates that all data must be removed from the servers. Instead, the data manager can try to procure a continued use license when the agency will no longer continue purchasing datasets.
- If faced with data removal due to end of contract terms, then the data manager must set up a contingency plan for reports/queries based on data that may face a "stop use". The data manager must work with the data provider to understand what happens to prior analyses, reports, and documents that used the data (e.g., will prior documents have to be destroyed?).

Even when data is free for government use, such as through another agency, the data manager must understand use rules and sharing restrictions. Also, it is desirable to create an agreement regarding the frequency and method of data updates.

The data manager is responsible for creating the long-term **data operations and maintenance plan**. This involves:

- Determining whether the data repository will be managed and maintained internally or by a corporate service center.
    - It's more expensive to manage and maintain internally due to the resources required, but there is control over staff responsiveness and project lifecycle.
    - A service center could cost less overall, removes the hassle of maintaining machines/servers, and ensures DBA depth, but there is less control over the staff responsiveness or project lifecycle. If the service center is outsourced to contractors, this may introduce need-to-know issues with the datasets.
- Determining whether there will be a static data schema, into which data formats will be forced to fit; or whether there will be a dynamic data schema that changes annually to fit the current data formats. The NISAC data manager recommends a static data schema.

- Keeping track of the lifecycle of data sources (e.g., when licenses need to be renewed).
- For each dataset, documenting where it is stored, its format, the version, the need-to-know rules, and the current access lists.
    - When comparing results, it is important to know the version of the data being used by all parties. For example, NISAC results were reporting conflicting information compared to another lab. The information conflict was caused by NISAC and the other lab using different versions (years) of a dataset.
- Budgeting for fixing data format issues annually. The NISAC data manager estimates that it cost $50K annually to deal with data format changes.
    - The data itself may have changed.
    - Fields will change name or type.
    - New fields will appear.
    - Existing fields will disappear.
- Developing software requirements that require software to be designed to be flexible to the data format changes. This can be achieved by abstracting the data layer.

The data manager must be cognizant of the **security requirements** for each data set. The data provider's data protection rules will be included in the contract terms. However, since the data repository will be storing multiple datasets, the combination of certain datasets can escalate the classification level of the entire repository. This is also a concern for any analysis or document that includes this combination of datasets. Work with a derivative classifier (DC) to determine in advance which combinations of datasets are of concern. It is important to perform this determination before adding a new dataset to the repository to avoid accidentally creating a classified repository, which would then require the entire server to be sanitized. The DC does not have to be a member of the data management team, but there should be a DC appointed to support the team.

The data manager must create a **data recovery plan**. On the NISAC program, every 6 months the data management team tested that the databases could be restored from backups.

The data manager must understand the **data retention and retirement rules**. What is the corporate policy on data retention? When can the data be removed from servers or stop being archived? What happens when corporate policy on retention conflicts with the contract terms that limit how long the data can legally be used? What happens if there is no longer software that can even run old models and datasets?

4. ANALYSIS AND RESULTS

A deep understanding of the data sources, data sets, and data collection methods are required to correctly perform an analysis and interpret results. Many times, analysis is performed using data sources/sets based on availability and without this deep understanding. Building a tool or dashboard on top of the data sets allows analysts to expose useful information to others while at the same time reducing the probability of erroneous interpretations by limiting access to raw data.

4.1. User Roles

User roles based on analytic expertise is a good method for controlling access to data sets and results. Advanced analysts would be those trusted to learn or who already understand the nuances of the data sets. They are the information producers. They will perform analyses using the raw data sets. They will also typically generate the reports and create (or inform the creation of) the tools and dashboards used by others.
At the other extreme are the view-only analysts. They are information consumers. They are not allowed to have access to the raw data sets. Instead, they can access properly vetted results through reports and tools. This does not mean they are limited to static graphs. They can still interact with dynamic visualizations if the visualizations have been vetted by the advanced users.

An implementation detail is whether all of the user roles will use the same or different tools. In our experience, this is a function of the complexity of the data and analysis. Advanced users may require the freedom to utilize multiple tools and models to perform their analysis. Once the analysis is complete, a report, tool, or dashboard is generated to communicate the results to the view-only users (**Error! Reference source not found.**).
OCIA's current plan to use Tableau as a common analytic platform and Tableau Reader to deliver results effectively implements the idea of user roles.
A single tool for all users makes sense in cases where:

- a custom model or algorithm must be created for the analysis,
- the model or data is proprietary (requires access controls), or
- custom visualizations would benefit all users.

Access controls based on user roles controls which functionality a user can access in the tool. As an example workflow (**Error! Reference source not found.**), the advanced users can perform analyses by selecting data inputs and modifying model parameters. The tool provides the ability to view analytic results. The advanced users may perform tens of analytic runs, but they may only select 1-2 results to "publish" or release to the other user roles. A view-only user can only see the released analytic results and cannot perform any analyses themselves. The data inputs for each analytic result is locked and cannot be modified. The advanced and view-only users see the same results visualizations. The disadvantages of custom software are planning for the resources to support maintenance, future upgrades, and improvements and maintaining developers to perform these tasks.

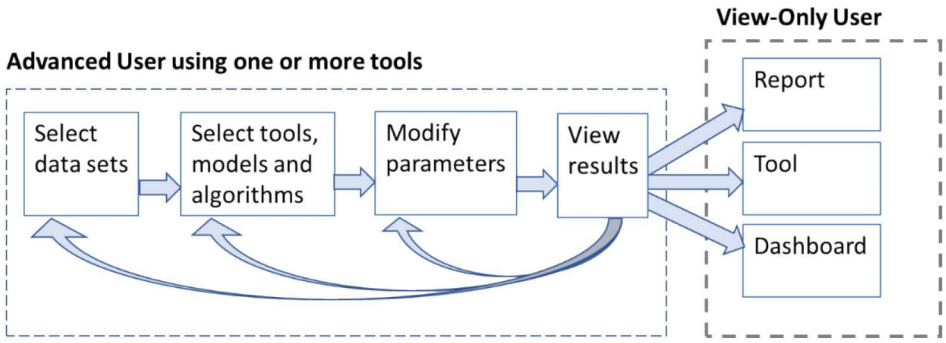

Figure 1. Workflow where the advanced and view-only users use different toolsets.

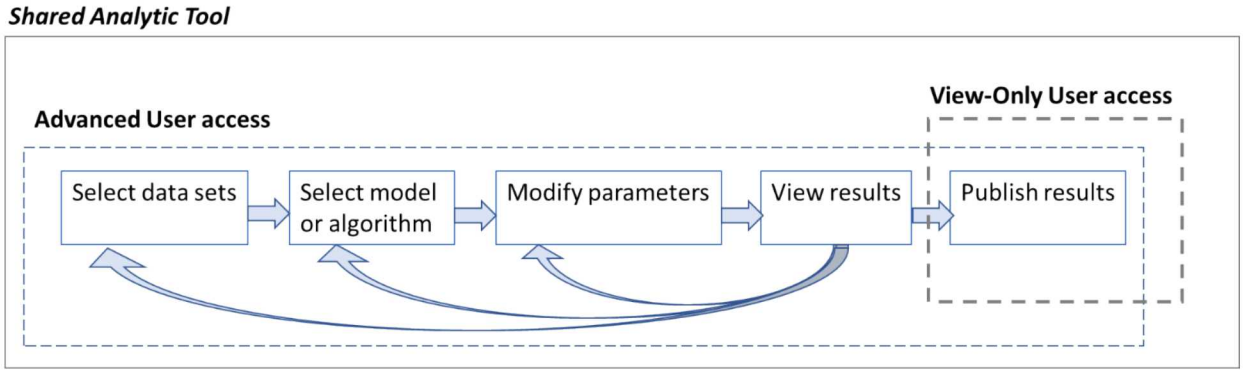

Figure 2. Workflow where advanced and view-only users use the same analytic tool.

4.2. Methodology and Assumptions

It is important to document the methodology and assumptions used in the analysis and production of results and visualizations.

4.2.1. Methodology

The body of this section should include:

- a description of the approach taken to produce any results displayed in the model and
- a description of the data sets used.

For example, this could include a mathematical formulation, a description of any software used, methods used to collect and process data, and any pre- or post-processing that impacts the results.

Examples:

- The model used is a mixed integer linear program, implemented using Python. The equations are as follows: …
- This is a simulation using the commercially available software FlexSim. We use data from the Army Corps of Engineers and then… etc.
- The Surface Transportation Board Waybill does not explicitly identify all containerized imports. Rather, we use the counties associated with each port as a surrogate to estimate what STB Waybill records might be imports. [5]

4.2.2. Assumptions

Include any assumptions that must be made:

- to aggregate the data,
- to simplify a model, or
- to interpret results.

This helps prevent misuse of the model and ensures that all stakeholders understand the limitations of the results produced.

Examples:

- Data is aggregated at the county-level, so the analysis will not capture details at a particular location.
- Simulations … were performed at the county level for one county per climate division overlying the High Plains Aquifer within Kansas and Nebraska (14 counties total) … This approach assumes that the outcome of crop yield simulations performed in one county represents the likely outcome for all counties within the same climate division. [4]
- We assume that all attackers have perfect knowledge of the system.
- It is assumed that additional supplies can be purchased, but at an additional cost.

5. EXAMPLES

5.1. Data Repository: Model and Data Inventory (MoDI)

The Modeling and Data Working Group (MDWG) was appointed by the Emergency Support Function Leadership Group (ESFLG) to create the Modeling and Data Inventory* (MoDI). The effort appears to have begun around 2013 or earlier. The goal was to build a "phonebook" of models, tools and datasets, "with the biggest driving force being to know what resources existed and to avoid JIT [just in time] discovery during a response." † A major challenge of the effort was that all the models, tools and datasets are owned or developed by other agencies. Resources were discovered through interviews with resource owners and obtaining additional contacts.

The MoDI was last updated on 9/19/2016. In 2018, the MDWG is working on a content update. The FEMA-MDWG@fema.dhs.gov mailing list appears to be the primary means for contacting the known technical points of contact (POCs) for each resource. To be included in the MoDI inventory, the resource must be operational and have at least one current agency-level user. Resources will be removed when they are no longer operational or no longer in use.

Each resource is tagged with an extensive set of metatags, when information is available. Each resource has a detailed information page with a standard format. POCs are listed to allow Federal Emergency Management Agency (FEMA) to be proactive and request access in advance. At the time of this writing, MoDI includes 320 resources. To make it easier to find relevant resources, the inventory can be filtered by:

- Hazard
- Emergency Support Function
- Owner (Agency)
- Recovery Support Functions
- Keyword

Resources are further categorized by function (resources can fall under multiple functions):

- Raw data
- Event characterization
- Situational Awareness
- Consequence models
- Impact Estimates
- Decision Support Tools
- Mission-specific requirements

MoDI also provides a network graph view of the entire inventory, displaying which models and tools use which data sets and model/tool outputs.

---

* https://gis.fema.gov/Model-and-Data-Inventory/index.html

† https://data.femadata.com/MDWG/05_May_MoDI_UserFeedback/May%202018%20MDWG%20MeetingSummary_MoDI%20User%20Feedback.pdf

5.2. Data Repository: Homeland Security Enterprise (HSE) Geospatial Concept of Operations (GeoCONOPS)

The Homeland Security Enterprise (HSE) Geospatial Concept of Operations (GeoCONOPS) is managed by the DHS Geospatial Management Office (GMO), under the Office of the Chief Information Officer. The goal of the GeoCONOPs program is to:[‡]

- Identify stakeholders and any capabilities, products, and datasets that they currently provide
- List technologies that are proven to address mission requirements
- Reduce duplication of effort
- Improve communication across entities at all levels (federal, state, local) providing geospatial support
- Ensure that information reaches front-line mission owners

The Geospatial Interagency Oversight Team (GIOT) guides the development of the GeoCONOPS program and meets annually to review the program and to identify gaps. At the time of this writing, the GeoCONOPS includes ~200 resources. Summary data is provided for each resource:

- A brief overview
- Category
- Unclassified URL
- Secret URL
- Top Secret URL

Information is organized into five main areas:

- Stakeholders – Lists stakeholders that have contributed to this program.
- GeoData & Products – Search or filter (by category and/or agency) all resources that provide geospatial information or products created to aid Homeland Security Missions and Support National Preparedness.
- Capabilities – Search or filter (by agency) for capabilities that support mission planning, rehearsal, and execution.
- Tradecraft – Search or filter (by agency) for training, operating procedures, capability assessment tools, templates, and other resources related to tradecraft.
- Best Practices – A list of methodologies, techniques, and procedures, typically derived from case studies.

[‡] https://cms.geoplatform.gov/geoconops/

5.3. Analytic Framework: Economic Scenario Analysis

SNL National Infrastructure Simulation and Analysis Center (NISAC) has attempted to develop a structured multi-hazard, multi-infrastructure scenario process as part of its mandate to develop model-based risk analysis approaches. NISAC economic analysts have published their own economic scenario analysis process [2], which will be summarized below. While this process is specific to economic analyses, the workflow is useful for any analysis involving infrastructure dependencies and multi-level geographic resolution (local, regional, national). An example economic analysis of Hurricane Katrina is provided in the paper.

5.3.1. Step 1: Translating scenario condition to direct impacts to critical infrastructure

The goal of this step is to estimate the loss of critical infrastructure performance in terms of outage duration and over time. This is often defined for each infrastructure asset in the region of interest. Examples provided in the paper:

> *Node-level physical damage to a communications network does not likely result in widespread loss of functionality, due to the redundant nature of these networks' structures; a cyber-attack, however, could rapidly cause widespread damage. Damage to a rail transportation rail yard will cause more damage to rail shipments than loss of a rail bridge over a major river, due to the inability to process train shipments.* [2]

First, identify the type of impact to critical infrastructure:

1. *Disruptions to networks (e.g., assets [nodes] or connections [edges] within transportation systems, energy, communications, or information technology);*
2. *Disruptions across networks, where the network is the medium for the attack (e.g., pandemics, cyber-attacks); and*
3. *Geospatial disruptions, where broad areas containing one or more of the critical infrastructure experience significant physical damage (e.g., hurricanes, floods).* [2]

Second, identify the scenario disruption dynamics:

1. *Where and at what level the scenario impacts the infrastructure system,*
2. *The topology of the network itself,*
3. *The speed at which scenario impacts cascade through the infrastructures, and*
4. *The effort required to restore operations.* [2]

To estimate item 3 above, they have developed a table (**Error! Reference source not found.**) to help determine the speed of direct economic impacts between any two infrastructures that are dependent. The row infrastructure is dependent on the column infrastructure.

Infrastructure	Agriculture & Food	Banking & Finance	Chemicals & Hazardous Materials	Commercial Facilities	Critical Manufacturing	Dams	Defense Industrial Base	Emergency Services	Energy / Electric Power	Energy / Natural Gas	Energy / Petroleum	Government Facilities	Healthcare & Public Health	Information Technology	Nuclear Facilities	Postal & Shipping	Telecommunications	Transportation / Air	Transportation / Highway	Transportation / Maritime	Transportation / Rail	Water
Agriculture & Food	mi	d	mo			d			mi	h	d					d		d	d	d	d	d
Banking & Finance		i							h					mi			i		h			
Chemicals & Hazardous Materials		d	mi						mi	h	h			h			d		h	h	h	h
Commercial Facilities	h	d	d	h	h			h	mi	mi				d		d	i	d	h	d	d	h
Critical Manufacturing		d	d	d	h			h	h	h				d		d	h		h	d	d	
Dams						i													mi	mi		
Defense Industrial Base		d					h											d	h		d	
Emergency Services								h	mi		d		h				mi	d	mi	d		mi
Energy / Electric Power		d				mi			i	h	d			mi	mi		mi		h		d	h
Energy / Natural Gas		d							mi	mi	d			h			h			h		
Energy / Petroleum		d							mi		mi			h			h		h	h	d	h
Government Facilities		d		h			mo	h	mi	mi	d	h		h		h	h		h		d	h
Healthcare & Public Health	d	d	d	d	mo			h	mi	mi	h		h	h	h	h	mi	d	mi			h
Information Technology		d							mi		h			mi			h					d
Nuclear Facilities					mo			h	mi					d	d		mi		h		d	h
Postal & Shipping		d														h						
Telecommunications		d			mo				d		d			h			i	d	h			
Transportation / Air		h		d	d				mi		h			h			mi	mi	h			
Transportation / Highway				d	d						d	mo							mi	mi	mi	
Transportation / Maritime		d		d	mo	h					d	mo		d			mi		h	mi	mi	
Transportation / Rail				d	d						d			d			mi		h	mi	mi	
Water		d	d		mo	h			mi					h					d		d	mi

Figure 3. Speed of infrastructure dependencies for select critical infrastructure. Legend: i = instantaneous dependence, mi = minutes, h = hours, d = days, mo = months. [2]

5.3.2. Step 2: Estimating infrastructure interdependencies

Building on step 1, the goal of this step is to generate the complete set of infrastructure impacts and to estimate the additive duration of their impacts. They depend on discussions between infrastructure subject-matter experts and modeling to determine the full scope and timing of infrastructure interdependencies. They define infrastructure interdependencies in terms of the four types of connections suggested by Rinaldi et al. [4] and Brown [1] and the speed with which these connections transfer the disruption:

1. *geographic connections or dependencies, where two or more infrastructure systems are co-located within the disruption zone;*
2. *physical connections or dependencies, where one system is directly connected to the other (a communications system connected to an energy system);*

3. *cyber dependencies, where one infrastructure uses information from another via information technologies; and*
4. *logical dependencies, where an infrastructure system takes actions based on non-proximal, non-physical, non-cyber dependencies (a firm buying foreign goods based on the availability of one or more ports of entry).* [2]

They have developed a table (**Error! Reference source not found.**) that displays the connections or dependencies for select critical infrastructures. An 'X' in a row denotes that the row infrastructure is dependent on the column infrastructure.

Infrastructure	Agriculture & Food	Banking & Finance	Chemicals & Hazardous Materials	Commercial Facilities	Critical Manufacturing	Dams	Defense Industrial Base	Emergency Services	Energy / Electric Power	Energy / Natural Gas	Energy / Petroleum	Government Facilities	Healthcare & Public Health	Information Technology	Nuclear Facilities	Postal & Shipping	Telecommunications	Transportation / Air	Transportation / Highway	Transportation / Maritime	Transportation / Rail	Water
Agriculture & Food	x	x	x			x			x	x	x					x		x	x	x	x	x
Banking & Finance		x							x					x			x		x			
Chemicals & Hazardous Materials		x	x						x	x	x			x			x		x	x	x	x
Commercial Facilities	x	x	x	x	x			x	x	x				x		x	x	x	x	x	x	x
Critical Manufacturing		x	x	x	x			x	x	x				x		x	x		x	x	x	
Dams						x													x	x		
Defense Industrial Base		x					x											x	x		x	
Emergency Services								x	x		x		x				x	x	x	x		x
Energy / Electric Power		x				x			x	x	x			x	x		x		x		x	x
Energy / Natural Gas		x							x	x	x			x			x			x		
Energy / Petroleum		x							x		x			x			x		x	x	x	x
Government Facilities		x		x			x	x	x	x	x	x		x		x	x		x		x	x
Healthcare & Public Health	x	x	x	x	x			x	x	x	x		x	x	x	x	x	x	x			x
Information Technology		x							x		x			x			x					x
Nuclear Facilities					x			x	x					x	x		x		x		x	x

Figure 4. Select critical infrastructure dependencies. [2]

5.3.3. *Step 3: Estimating direct impacts from scenario and infrastructure impacts*

The goal of this step is to estimate the direct impacts. This is done by taking the set-theory union of all infrastructure direct effects (over time and regions) from the subject-matter expert discussions and translating them into direct impacts.

5.3.4. *Step 4: Estimating total (direct plus indirect) impacts*

The goal of this step is to identify and add indirect impacts to the direct impacts estimate. The appropriate secondary and tertiary indirect impacts will have to be determined for the analysis domain. In terms of economic analysis, secondary indirect impacts are the impacts "to the economic firms involved in commerce with the directly

impacted firms" [2], and the tertiary indirect impacts "are caused by the loss of income to employees of impacted firms." [2]

For regional scenarios (e.g., hurricanes, earthquakes), economic impact is also "divided spatially into those that occur within the physically damaged areas and those that occur outside of it. These estimates are made at the state, county, and zip code levels."

5.3.5. Step 5: Quantifying uncertainties and validating results

The majority of data available are "point estimates" of baseline conditions of impact. Additional effort is made to set range estimates for the impacts. Uncertainties in impact estimates are caused by uncertainties in "scenario effects, infrastructure direct and interdependency effects, direct and indirect economic effects, and data and model errors." Whenever possible, they attempt to validate their results against estimates made by others or against "post-event "on the ground" data of actual impacts." Unfortunately, there often isn't data available to perform these comparisons.

5.3.6. Step 6: Reporting results

Based on the scenario and infrastructures in the analysis, it is helpful to pre-determine the standard set of results that will be. A checklist of results will (1) save time when analysts don't have to brainstorm an ad hoc list of metrics and results for every request and (2) ensure some consistency in analyses and reports among multiple analysts. The OCIA use case development framework is a step in this direction.

NISAC reports the following scenario and infrastructure impacts:

- Estimates of displaced civilians
- Morbidity and mortality
- Temporary and permanent losses of infrastructure functionality
- Transportation
- Electric power
- Telecommunications
- Ports of entry

As an example, the economic analysts report additional economic impacts.

1. Macroeconomic impacts
    a. Lost Gross Domestic Product (GDP)
    b. Lost employment
    c. Lost trade
    d. Changes in prices
2. Mesoeconomic impacts
    a. Impacted supply chains
    b. Impacted commodity shipments

3. Microeconomic impacts
    a. Numbers of businesses impacted by
        i. Location
        ii. Economic sector
        iii. Type of disruption
        iv. Duration

5.4. DHS Emergency Support Functions (ESF) Decision Support Tool (DST)

When the set of analytic questions are known and can be standardized, an analytic framework can be established to address those questions. The analytic framework can then be turned into a tool. Depending on the expertise of the user base, a wizard-like tool can walk the analysts through the analytic process.

An example of an analysis "wizard" is the DHS Emergency Support Functions (ESF) Decision Support Tool (DST). SNL led the development of the DST for FEMA. We are not aware of the current state or status of the tool. According to tool's documentation, the DST "integrates the planning factors supporting ESF #3 (Public Works & Engineering), #6 (Mass Care), #7 (Logistics), and #12 (Energy) into a common platform. The user interface provides a common set of inputs to support calculations for any combination of these ESFs, ensuring that the assumptions and calculations are consistent across the ESFs." This common platform also ensures some consistency in analyses performed among multiple analysts.

The tool is distributed with the Hazus infrastructure data, static FEMA Distribution Center Commodities, Region X planning factors, and 2014 U.S. Census population data pre-loaded. To run all the ESFs, the user is required to obtain and upload HIFLD Open and HIFLD Secure (our demo version still refers to them as HSIP Freedom or HSIP Gold) and EAGLE-I customer outages data. Once all necessary datasets are uploaded in the DST, the data layer is abstracted away from the user's view, and the tool handles pulling the relevant data for the selected ESF and event characteristics.

Other datasets specific to the hazard type (hurricanes, earthquakes, IND) are needed as references. The tool does not connect to any external tools or models. It is assumed that any modeling or analysis to define the characteristics and impacts of the hazard has already been completed.

The analysis begins by selecting a hazard and one or more ESFs. The user is guided step-by-step through inputs to characterize the event, such as location of incident, damage extend, population seeking shelter, resource distribution, power outages, and damaged infrastructure/generator needs. Default values are provided where applicable, with the ability to modify some values. Calculations are performed by the tool. Pre-defined results are organized by ESF. While some graphs are static, others are interactive, such as maps and generator transportation graphs. Values and assumptions made on the inputs pages are displayed in the result charts and graphs. Calculation assumptions are displayed at the bottom of each ESF results page.

REFERENCES

1. Brown, T. (2008). Infrastructure Dependency Indicators. *Wiley Handbook of Science and Technology for Homeland Security*.
2. Ehlen, M. A., & Vargas, V. N. (2013). Multi-hazard, multi-infrastructure, economic scenario analysis. *Environment Systems & Decisions*, *33*(1), 60-75.
3. Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems*, *21*(6), 11-25.
4. Tidwell, V. C., Vargas, V. N., Jones, S. M., Dealy, B. C., Shaneyfelt, C., Smith, B. J., & Moreland, B. D. (2016). *Analysis of high plains resource risk and economic impacts* (No. SAND2016--3405). Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States).
5. Wang, H., Gearhart, J., Jones, K., Frazier, C., Nozick, L., Levine, B., & Jones, D. (2016). Estimation of an origin–destination table for US imports of waterborne containerized freight. *Transportation Research Record: Journal of the Transportation Research Board*, (2548), 35-42.

DISTRIBUTION

3 Department of Homeland Security Office of Cyber and Infrastructure Analysis
Attn: K. Kilby, R. Hanson, C. Zapata
NPPD/OCIA
245 Murray Lane, M/S 0390
Washington, DC 20528-0290

1	MS1188	Alisa Bandlow	08831
1	MS1128	Stephen Kleban	08832
1	MS 9159	Noel Nachtigal	08700
1	MS0899	Technical Library	9536 (electronic copy)

Sandia National Laboratories