D2dprov: Vision 2025

A transdisciplinary science, technology, and policy synthesis on data-driven, science-informed resilience planning for 2025 and beyond

---

Author: Brian Wee (ORCID: 0000-0002-0038-9381)

Organizational affiliations: Founder and Managing Director, Massive Connections, LLC
Senior Advisor for Data, Science, and Policy, Neptune and Company, Inc.

Date released: 2019-02-21

Version: 1.2

Citation: Wee, B. (2019). D2dprov: Vision 2025. A transdisciplinary science, technology, and policy synthesis for data-driven, science-informed resilience planning for 2025 and beyond. Washington, DC. https://doi.org/10.6084/m9.figshare.7591238

Team members: William Teng (ORCID: 0000-0001-5642-6359)
Nancy Hoebelheinrich (ORCID: 0000-0002-6797-7903)
Tom Narock (ORCID: 000-0002-9785-4496)
Arif Albayrak

ESIP Lab project title: Enabling the encoding and visualization of provenance metadata for better discovery and understanding of climate resilience strategies for agriculture-related decision-making.

Acknowledgment: This work is based on team collaboration tools provided by the ESIP Lab with support from the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA) and the United States Geologic Survey (USGS).

Citation BibTeX: @techreport{Wee2019,address = {Washington, DC},author = {Wee, Brian},doi = {10.6084/m9.figshare.7591238},institution = {Massive Connections},title = {{D2dprov: Vision 2025. A transdisciplinary science, technology, and policy synthesis for data-driven, science-informed resilience planning for 2025 and beyond.}},year = {2019}}

Table of Contents

Table of Figures

1.0 Scope of the d2dprov project

Imagine that a coastal county finds itself at the start of a planning cycle to make the county "climate ready". These planners would greatly benefit from a fully traceable solution already implemented by another locality faced with similar socio-environmental challenges that can be used to bootstrap the local planning process. Just as one could download open-source scientific code from GitHub, imagine the benefits of a fully traceable resilience workflow (i.e. "recipes" for resilience planning) that is captured as a knowledge graph and shared as an open resource. That knowledge graph captures digital artifacts, or pointers to digital artifacts, including data, models, journal articles, integrated scientific assessments, applicable building codes, economic projections, risk assessment frameworks, formal decision-analysis, stakeholder priorities, and other relevant artifacts. Because each artifact is assigned a unique identifier and semantically linked to other relevant bodies of knowledge, the planning team is empowered to follow lines of inquiries that may otherwise have been difficult if not for the possibility of knowledge graph traversals.

The d2dprov (short for "data to decisions provenance") project is aimed at assessing the technologies required to implement a "GitHub" for resilience planning as outlined above. Figure 1 provides a high-level overview of the project.

This document outlines the vision for data-driven, science-informed, traceable resilience planning with an outlook to the year 2025 and beyond. A separate document provides additional technical details: "D2dprov: Statement of Needs 2022. Technology and policy requirements to fulfill Vision 2025's proposed approach to data-driven, science-informed climate resilience decisions" (Wee, 2019a).

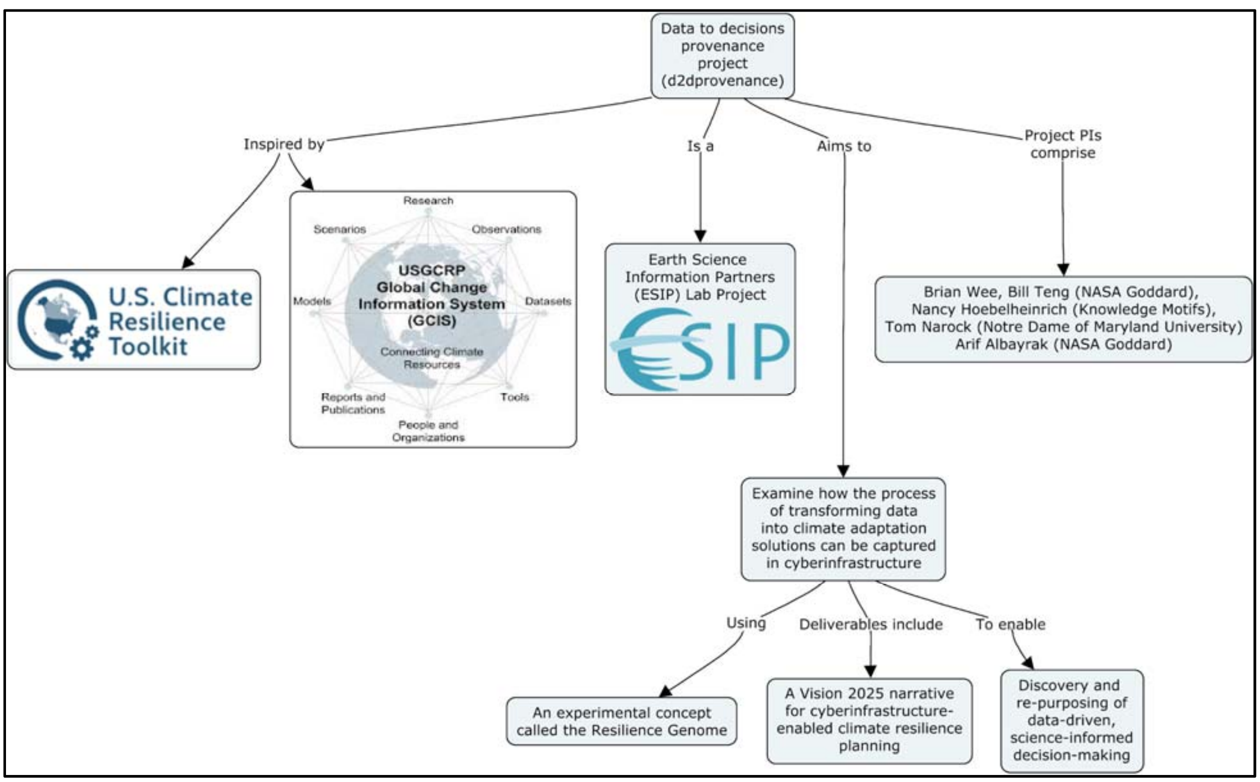

Figure 1: Scope of the d2dprov project.

The following table lists resources developed by the d2dprov team that are freely accessible. The "Reference" column reflects entries in the bibliography that resolve to web resources. URLs are not listed in this column to facilitate maintenance and evolution of this document.

Type	Title	Description	Reference
Report (this report)	D2dprov: Vision 2025 A transdisciplinary science, technology, and policy synthesis on data-driven, science-informed resilience planning for 2025 and beyond	Outlines the vision for data-driven, science-informed, traceable resilience planning with an outlook to the year 2025 and beyond.	Wee, 2019b
Report	D2dprov: Statement of Needs 2022. Technology and policy requirements to fulfill Vision 2025's proposed approach to data-driven, science-informed climate resilience decisions.	Outlines selected technological and policy components that will contribute to the fulfillment of the goals outlined in "Vision 2025".	Wee, 2019a
Python code, documentation, and presentation slides	Prototype parser for concept extraction using community ontologies and natural language processing	Parses Environmental Impact Statements (EIS) and Records of Decisions (ROD) using Natural Language Processing (NLP) to identify terms from environmental and resilience ontologies.	Narock, Wee, Hoebelheinrich, Albayrak, & Teng, 2019
Documentation and presentation slides	Experimental modeling of workshop planning processes to W3C PROV	Describes the outcome of an experiment to ascertain how well W3C PROV could be used to capture the provenance of a planning process that used a National Park Service developed scenario planning process.	Narock, Wee, Hoebelheinrich, Albayrak, & Teng, 2019
Presentation slides	Strategy for advanced query and discovery of context-relevant resilience documents	Outlines a high-level approach for using machine learning to acquire the content structure of existing documents in order to return a set of documents that match the user's query.	Narock, Wee, Hoebelheinrich, Albayrak, & Teng, 2019

2.0 Context

2.1 Resilience planning exemplar

The state of New Jersey's "Rebuild by Redesign – Hudson River Project" (henceforth referred to as the "Hudson River Project") is used as an exemplar to contextualize the discussion in this document. The Hudson River Project "convenes a mix of sectors - including government, business, non-profit, and community organizations - to gain a better understanding of how overlapping environmental and human-made vulnerabilities leave cities and regions at risk" (http://www.rebuildbydesign.org/about). As a response to Superstorm Sandy (2012), New Jersey received over US$200M in funding from the US Housing and Urban Development authority to implement flood mitigation measures for Hoboken, NJ, immediately across the river from downtown New York City. Figure 2 shows an overview of the area of concern for the flood mitigation planning. Figure 3 shows an example of a storm sewer system that may be incorporated as part of the flood mitigation strategy. While a storm sewer system is considered "gray infrastructure" (i.e. built infrastructure) for severe weather mitigation, cities often consider "green infrastructure" strategies. Figure 4 shows examples of green and gray infrastructure often considered for coastal resilience planning.

Figure 2: Hoboken, NJ (New Jersey Department of Environmental Protection, 2017).

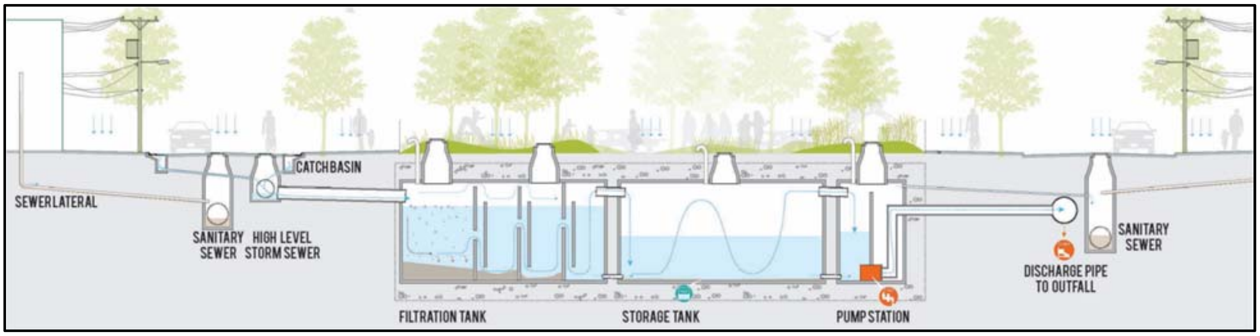

Figure 3: Storm sewer system for flood mitigation (New Jersey Department of Environmental Protection, 2017).

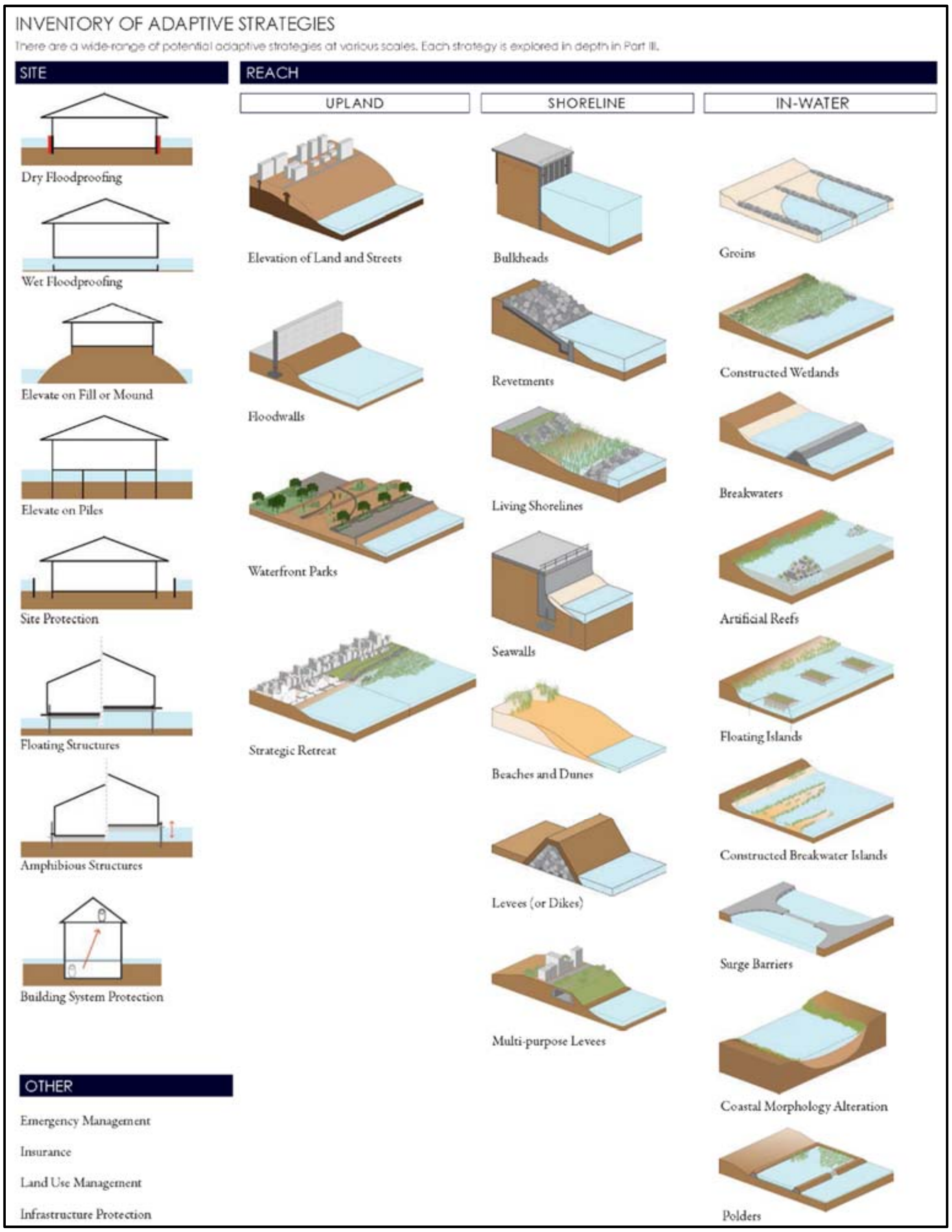

Figure 4: Examples of gray and green infrastructure strategies for coastal adaptation (Bloomberg, 2013).

2.2 Essential definitions

Presidential Policy Directive 8 (PPD-8: March 30, 2011, President Obama) called for the creation of a series of integrated national planning frameworks covering prevention, protection, mitigation, response, and recovery. This document focuses on "**mitigation**" activities as stipulated in PPD-8. The approaches outlined in this document are equally applicable to the other PPD-8 frameworks including prevention, protection, response, and recovery.

The terms "**mitigation plans**", "**resilience plans**", and "**adaptation plans**" are synonymous for the purpose of this document. It is tempting to envision a **plan** as a self-contained singular document, like a single PDF file. This is clearly not applicable because planning projects often result in a large collection of both machine- and human-readable digital artifacts, including narratives (e.g. journal publications, policy statements, administrative orders), data and model code. Plans for a resilience project may therefore be described as a collection of digital artifacts. Digital artifact access, storage, or representation are issues not addressed in this document.

A **resilience planning framework** provides a tractable means to conceptually organize digital artifacts associated with a resilience planning project. Such frameworks represent an abstracted causal model represented as logical steps that describe an ideal planning process. As such, a resilience planning framework may be used to enforce a temporal structure on the artifacts associated with a resilience plan. For this project, we focus on the US Climate Resilience Toolkit's "Steps to Resilience" planning framework (Figure 5). Examples of similar planning frameworks are provided in Section 4.2.

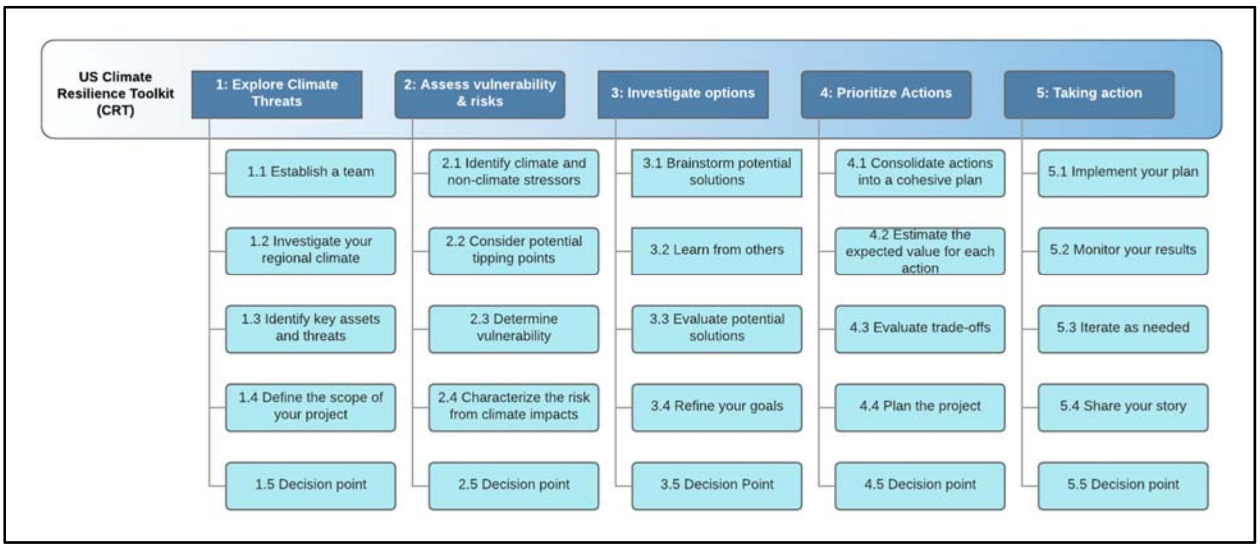

Figure 5: US Climate Resilience Toolkit's "Steps to Resilience" planning framework.

3.0 Extreme weather, climate change, and resilience

Local governments (e.g. state, tribal, county, city, local governments) increasingly carry the burden of implementing climate mitigation measures. This is evidenced by initiatives like the Rockefeller's 100 Resilient Cities and "block grants" offered by the US Department of Housing and Urban Development for local governments to upgrade civil infrastructure. The Hudson River Project (Section 2.1) was funded through a HUD block grant.

However, designing mitigation strategies in a manner responsive to **future forecasted environmental stresses** is challenging. Successful mitigation design depends on gaining access to a pool of human capital experienced in the complex choreography of climate resilience planning (see also Section 5.2.6 "State of resilience planning in the United States").

3.1 Costs of being unprepared

The Government Accountability Office (GAO), the investigative arm of Congress, started including climate change risk as a "high risk" to the nation starting 2013. In the GAO's bi-annual "High-risk Series" of report to Congress in 2017, the GAO noted that disaster relief appropriated by Congress for the period 2007 - 2013 was 46% higher than the period 2000 - 2006. The GAO re-iterated its recommendation to enhance climate resilience as a way to manage the nation's future fiscal exposure to the impacts of climate and weather impacts.

NOAA NCEI released a tool at https://www.ncdc.noaa.gov/billions/ (URL valid as of 2018-09-20T00:00:00Z) that depicts regularly updated snapshots of "weather and climate disaster events with losses exceeding $1 billion each across the United States" (Figure 6).

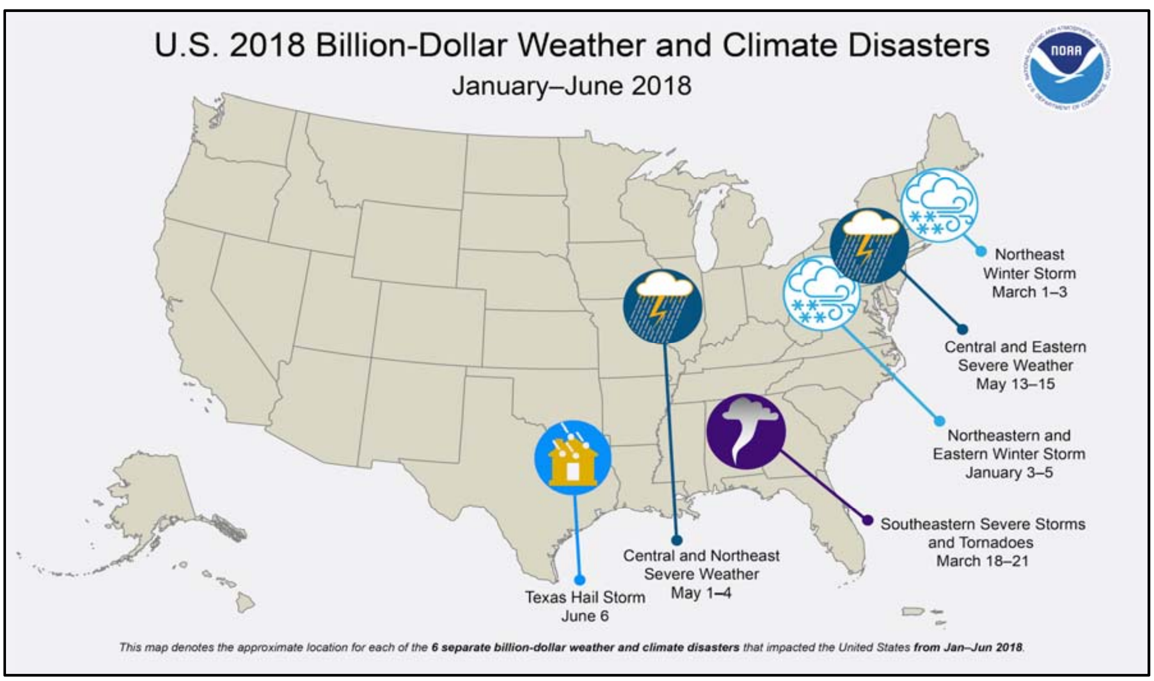

Figure 6: Billion-dollar weather and climate disasters January - June 2018.

3.2 US policy relevant to resilience planning

3.2.1 Presidential Policy Directive 8

Presidential Policy Directives (PPD) have the force of Presidential Executive Orders, but are sanctioned by the White House's National Security Council. Presidential Policy Directive 8 (PPD-8) titled "National Preparedness" was issued on March 30, 2011 by President Obama. The directive is:

> *"aimed at strengthening the security and resilience of the United States through systematic preparation for the threats that pose the greatest risk to the security of the Nation, including acts of terrorism, cyber attacks, pandemics, and catastrophic natural disasters."*

PPD-8 called for the creation of a series of integrated national planning frameworks covering prevention, protection, mitigation, response, and recovery.

PPD-8 uses an "all-hazards" approach to preparedness, and defines "mitigation" as "*capabilities necessary to reduce loss of life and property by lessening the impact of disasters*". This document was written primarily with the "mitigation" focus as it relates to **increasing resilience through adaptation by mitigating against hazards through the reduction of risk**.

The National Mitigation Framework established under PPD-8 describes seven core capabilities aimed at reducing loss of life and property from disasters. These core capabilities are: community resilience, long-term vulnerability reduction, operational coordination, planning, public information and warning, risk and disaster resilience assessment, and threats and hazards identification. Many of these capabilities align well with the overall objectives of the US Climate Resilience Toolkit (US CRT, administered by NOAA) established under Executive Order 13653.

3.2.2 Executive Order 13653

Executive Order (EO) 13653 "Preparing the United States for the Impacts of Climate Change" was enacted by President Obama's administration on 2013-11-06, and revoked by the following Presidential administration on 2017-03-18. EO-13653 § 1 stipulates (emphasis added):

> *…the Federal Government will continue to support scientific research, observational capabilities, and assessments necessary to improve our understanding of and response to climate change and its impacts on the Nation.*
>
> *The Federal Government must build on recent progress and pursue new strategies to improve the Nation's **preparedness and resilience**. In doing so, agencies should promote: (1) engaged and strong partnerships and information sharing at all levels of government; (2) **risk-informed decisionmaking and the tools to facilitate it**; (3) **adaptive learning, in which experiences serve as opportunities to inform and adjust future actions;** and (4) **preparedness planning**.*

Furthermore, EO-13653 § 4 stipulates (emphasis added):

> ***Providing Information, Data, and Tools for Climate Change Preparedness and Resilience.*** *(a) In support of Federal, regional, State, local, tribal, private-sector and nonprofit-sector efforts to prepare for the impacts of climate change…. and any other agencies as recommended by the Council established in section 6 of this order, shall, supported by* ***USGCRP****, work together to develop and provide* ***authoritative, easily accessible, usable, and timely data, information, and decision-support tools on climate preparedness and resilience.*** *(b) As part of the broader* ***open data policy****, CEQ and OSTP, in collaboration with OMB and consistent with Executive Order 13642 of May 9, 2013 (****Making Open and Machine Readable the New Default for Government Information****), shall oversee the establishment of a web-based portal on ''Data.gov'' and work with agencies on identifying, developing, and integrating* ***data and tools relevant to climate issues and decisionmaking****.*

The US Climate Resilience Toolkit (US CRT) and the US Global Change Research Program's (USGCRP) Global Change Information System (GCIS) serve to help fulfill the orders enacted in EO-13653.

3.2.3 PREPARE Act of 2017

The Preparedness and Risk Management for Extreme Weather Patterns Assuring Resilience and Effectiveness (PREPARE) Act of 2017 (House Resolution 4177, 115th Congress) was introduced by Congressman Cartwright (D-PA-8) calling for an interagency council to promote

> *"the development of innovative, actionable, and accessible Federal* ***extreme weather resilience, preparedness, and risk identification and management-related information, data, tools, and examples of successful actions at appropriate scales for decision makers****".*

In the explanatory report accompanying the legislative language, the US House of Representative's Committee on Transportation and Infrastructure referred to findings from NOAA that estimated the total cost of damage caused by extreme weather for the United States and findings from the Government Accountability Office (see above).

In the PREPARE bill, the term "extreme weather" includes:

> *observed or anticipated severe and unseasonable atmospheric conditions, including drought, wildfire, heavy precipitation, wave, high water, snowstorm, landslide, mudslide, hurricanes, tornadoes and other windstorms (including derechos), extreme heat, extreme cold, sustained temperatures or precipitation that deviate from historical averages*

In late 2018, the PREPARE Act of 2017 was subsumed into the Federal Aviation Authority's Reauthorization Act of 2018 (House Resolution 4, 115th Congress), and passed by the House of Representatives. Despite the House of Representatives passing the FAA Reauthorization 2018 bill out of its chamber, the final bill signed into law did not include the PREPARE Act. The bill awaits further action under the 116th Congress.

3.2.4 OPEN Government Data Act

The OPEN Government Data Act, passed into law on 2019-01-14, is part of the Foundations for Evidence-Based Policymaking Act of 2018. The OPEN Government Data Act amends the language under the US Code (44 USC § 3502) related to "Federal information policy" by defining the terms data, data asset, machine-readable, metadata, open Government data asset, open license, public data asset, and statistical laws. "Metadata" is defined as (emphasis added):

> *structural or descriptive information about data such as content, format, source, rights, accuracy,* ***provenance****, frequency, periodicity, granularity, publisher or responsible party, contact information, method of collection, and other descriptions*

Furthermore, the Act amends the responsibilities of federal agencies (44 USC § 3506). The Act requires that agencies coordinate with the Office of Management and Budget (OMB) to develop processes and procedures that:

> *require data collection mechanisms created on or after the date of the enactment of the OPEN Government Data Act to be available in an open format*

and to (emphasis added):

> ***facilitate collaboration with non-Government entities*** *(including businesses), researchers, and the public for the purpose of understanding* ***how data users value and use government data***

3.3 ESIP's role

ESIP has supported activities relevant to PPD-8, EO-13653, and the PREPARE Act of 2017 by way of facilitating the following:

- d2dprov project.
- ESIP clusters:
  - Data to Decisions (inactive as of pre-determined sunset date of 2017-12-31)
  - Agriculture and Climate
  - Disaster Lifecycle
  - Community Resilience
- The USGCRP's GCIS and US CRT have been represented at ESIP meetings and promulgated their vision and objectives through ESIP plenary sessions and breakout sessions. Figure 1 illustrates how the US CRT and GCIS formed the basis for many of the ideas presented in this document.
- In June 2017, Congressman Cartwright's office received official communication stating ESIP's sanction of the PREPARE Act of 2017, given the relevance of the proposed legislation to the ESIP constituency.

- ESIP was a signatory, together with many organizations, on a letter of support to Congress urging passage of the OPEN Government Data Act when it was first introduced for consideration.

4.0 Conceptual Frameworks

For the purposes of d2dprov, we subscribe to a distinction between (a) frameworks for assessing the socio-environmental landscape for a resilience planning project, and (b) frameworks for resilience planning.

Resilience planning frameworks are inherently process-oriented: they denote a series of steps that resilience planners are encouraged to follow. That process-orientation maps easily to workflow concepts, which the W3C PROV is designed to model. Such planning frameworks may also be used in **risk management** communities of practice, even if they are not described as such during the term "resilience planning". There appears to be little overlap between persons identifying themselves as belonging to the risk management community versus the resilience planning communities. Nevertheless, there are similarities in concepts and approaches adopted by both communities that are highly applicable to resilience planning.

Frameworks for assessing the socio-environmental landscape are often used extensively within a resilience planning project for "vulnerability / risk assessments". Such frameworks may be thought of as "embedded" within a resilience planning framework. Examples of both types of frameworks are provided below.

It is not hard to identify a small set of resilience planning frameworks that either (1) the community has identified as having gained widespread acceptance, and/or (2) funding agencies have identified as their preferred methodology and require that grantees utilize that methodology. For any framework within this library of frameworks, the steps within that framework should be cross-walked to steps in other frameworks, so as to render these planning frameworks interoperable. Ultimately, these cross-walk facilitates machine discovery of related resilience plans that have been created and described using different vocabularies. See Section 6.2 "Searching and using the Resilience Genome repository".

4.1 Frameworks for understanding the socio-environmental decision landscape for resilience planning

4.1.1 Socio-ecological framework (one of many variants)

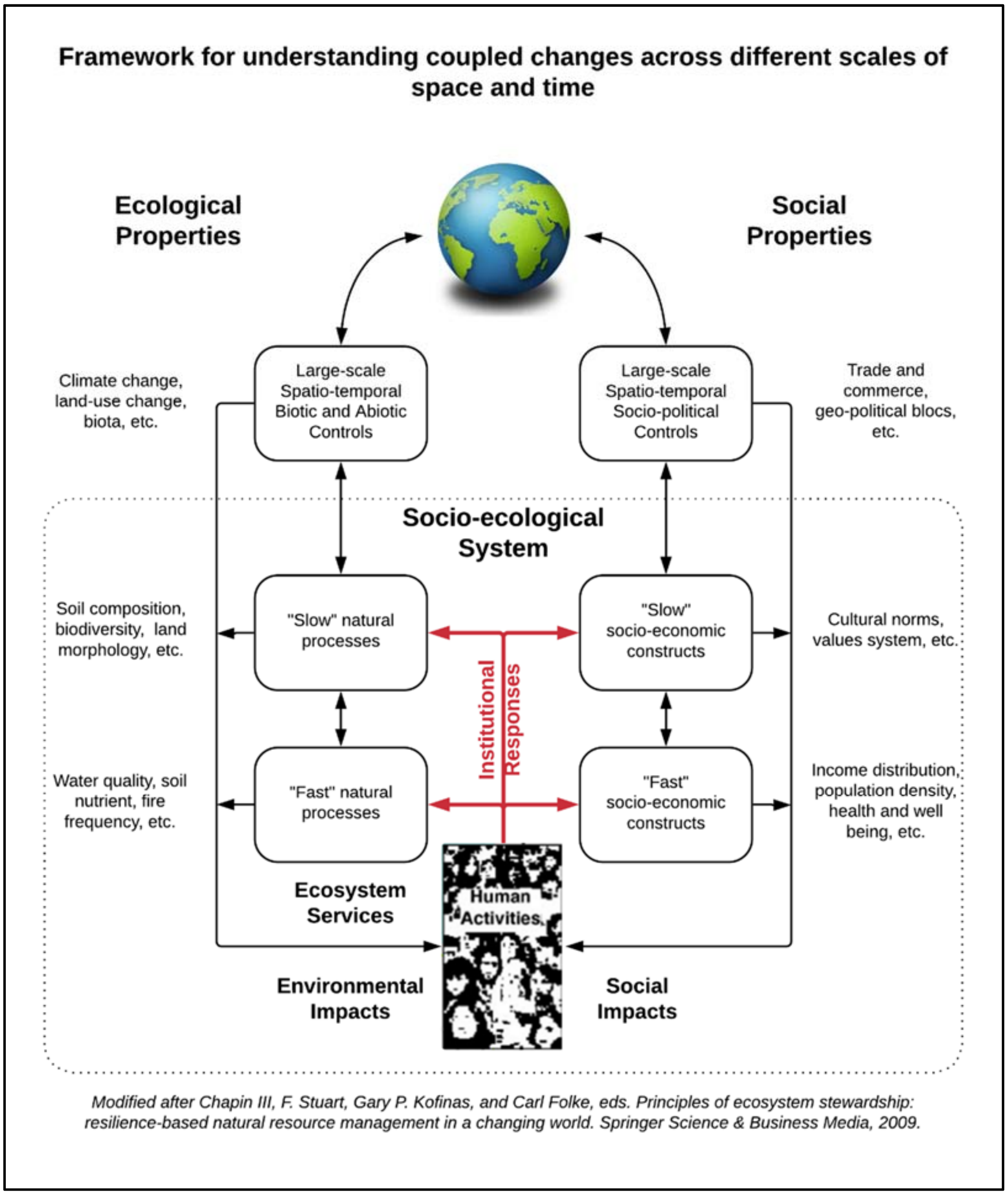

Figure 7: Socio-ecological framework

4.1.2 Drivers, pressures, state, impact and response model of intervention (DPSIR)

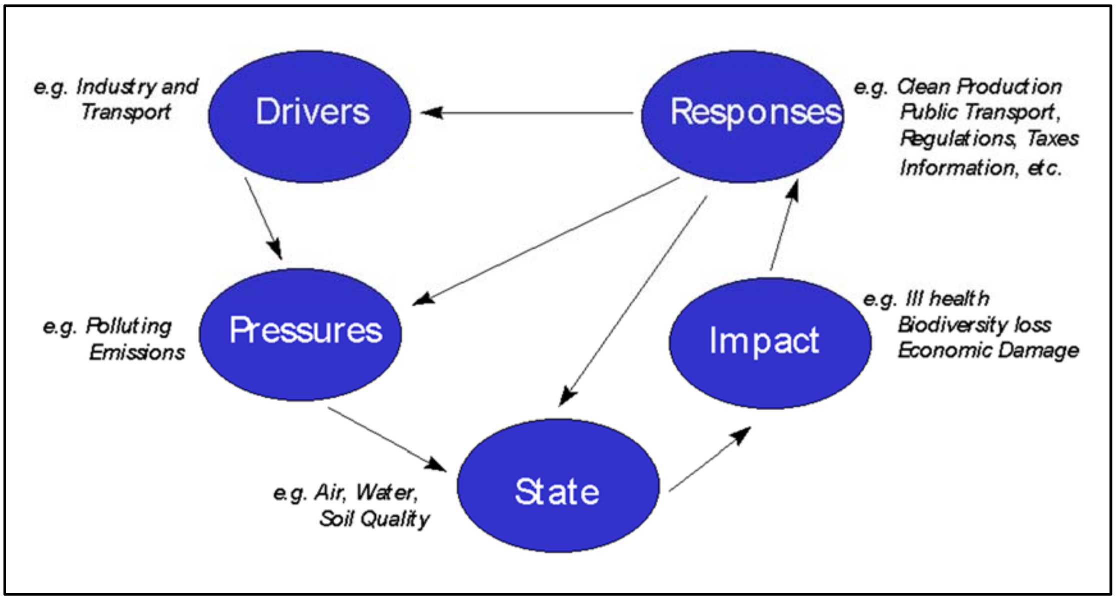

Figure 8: DPSIR framework (European Environment Agency)

4.1.3 Measurement framework for community resilience

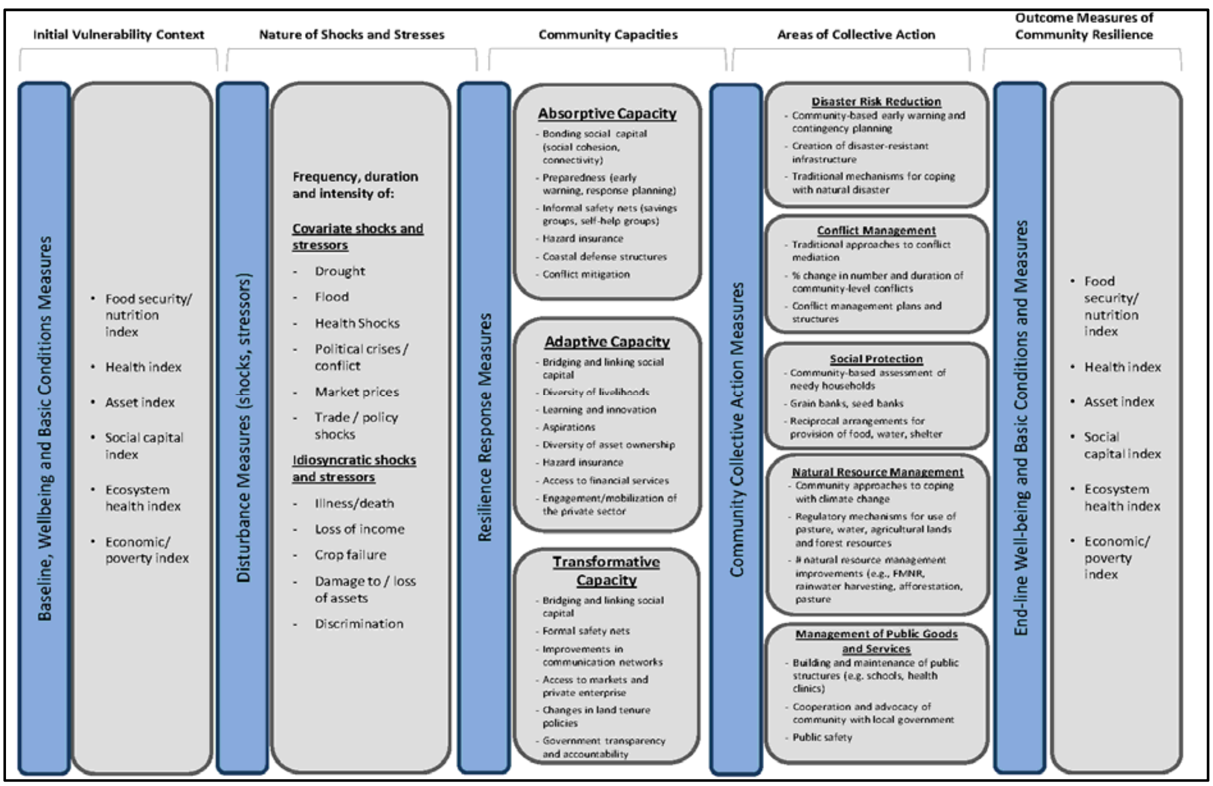

Figure 9: Measurement framework for community resilience (Mueller, Spangler, & Alexander, 2013)

4.2 Frameworks for resilience planning

4.2.1 US Climate Resilience Toolkit's "Steps to resilience" Framework mapped to Structured Decision Making

The figure below is a combination of Figure 5 and a formal decision analysis method called "Structured Decision Making" (SDM). SDM is described in Section 5.3 "Decision analysis for climate resilience".

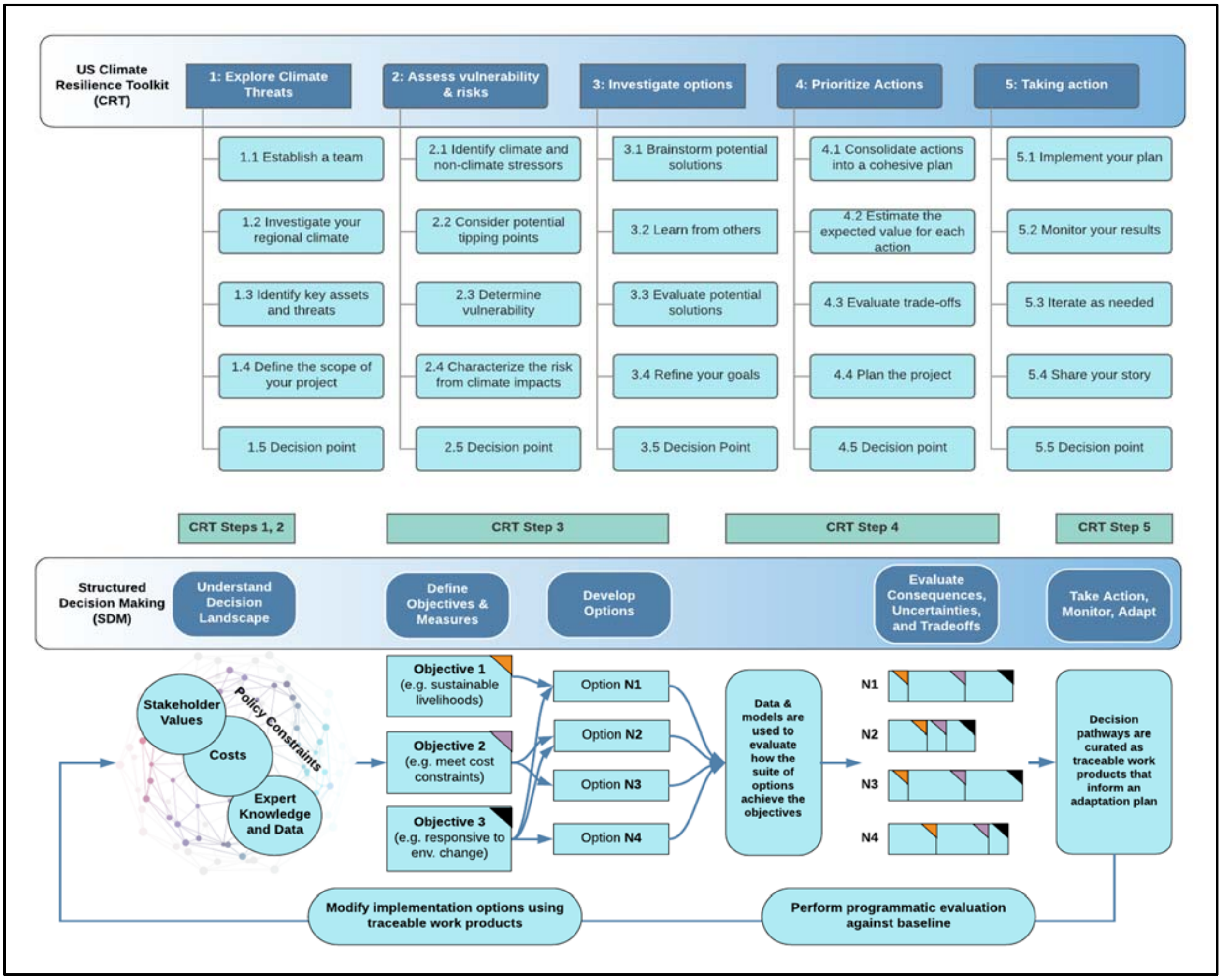

Figure 10: US Climate Resilience Toolkit planning protocol mapped to Structured Decision Making (Wee, 2019f)

4.2.2 USAID Climate-Resilient Development Framework

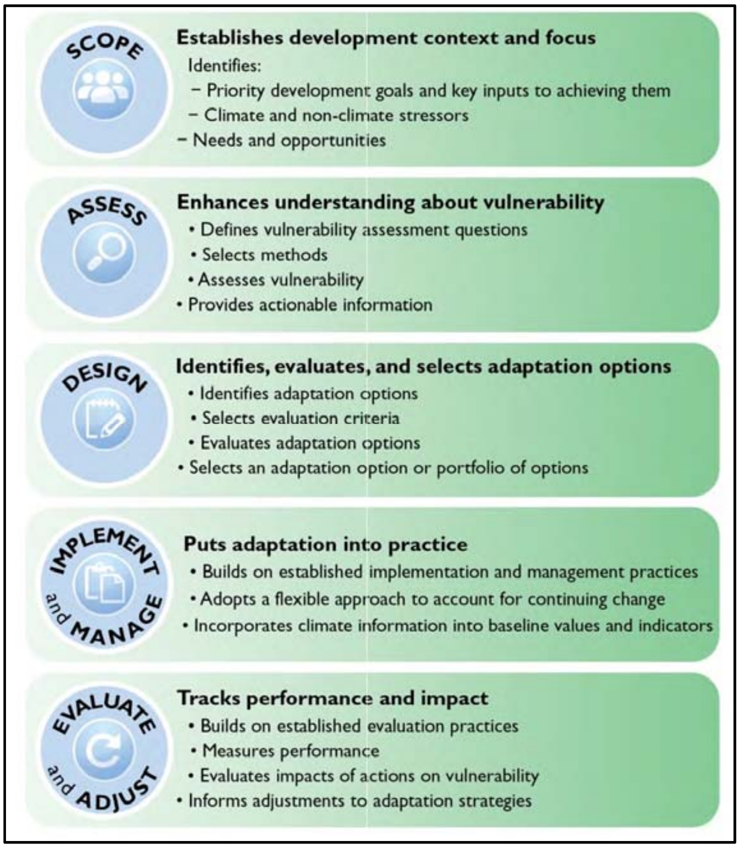

Figure 11: USAID climate-resilient development framework (*Climate-resilient development: A framework for understanding and addressing climate change*, 2014)

4.2.3 UKCIP risk framework

The UKCIP was formerly known as the "UK Climate Impacts Programme".

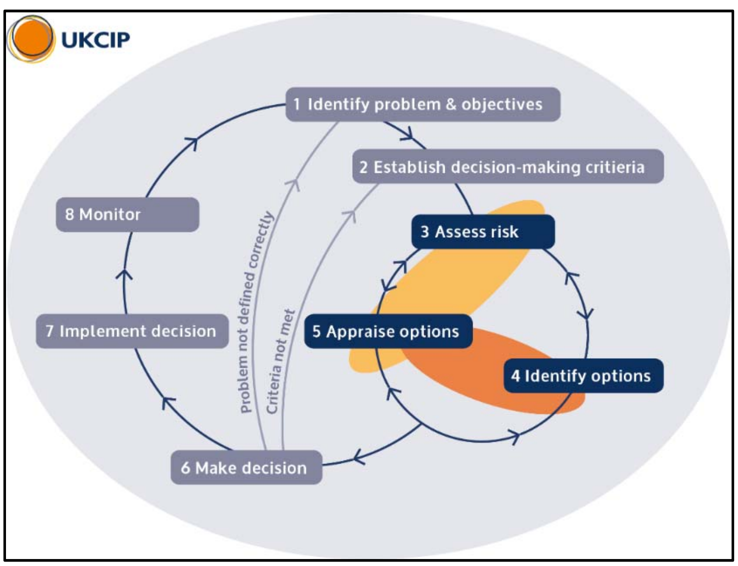

Figure 12: UKCIP risk framework ("UKCIP risk framework | UKCIP," n.d.)

4.2.4 National Park Service five-step scenario planning process

The National Park Service's five-step scenario planning process has been used in workshops "for developing multivariate climate change scenarios" (*Using Scenarios to Explore Climate Change: A Handbook for Practitioners*, 2013). A resilience workshop for beef agriculture stakeholders was convened by the University of Nebraska – Lincoln. Workshop organizers used this five-step planning process to engage stakeholders in a conversation about ways to adapt to the changing climate. D2dprov used a work product from that workshop to model the steps undertaken by workshop participants. The objective of this experiment was to ascertain how well W3C PROV could be used to capture the provenance of a planning process that used this five-step scenario planning process. The work products for this effort are documented in Narock et al., 2019.

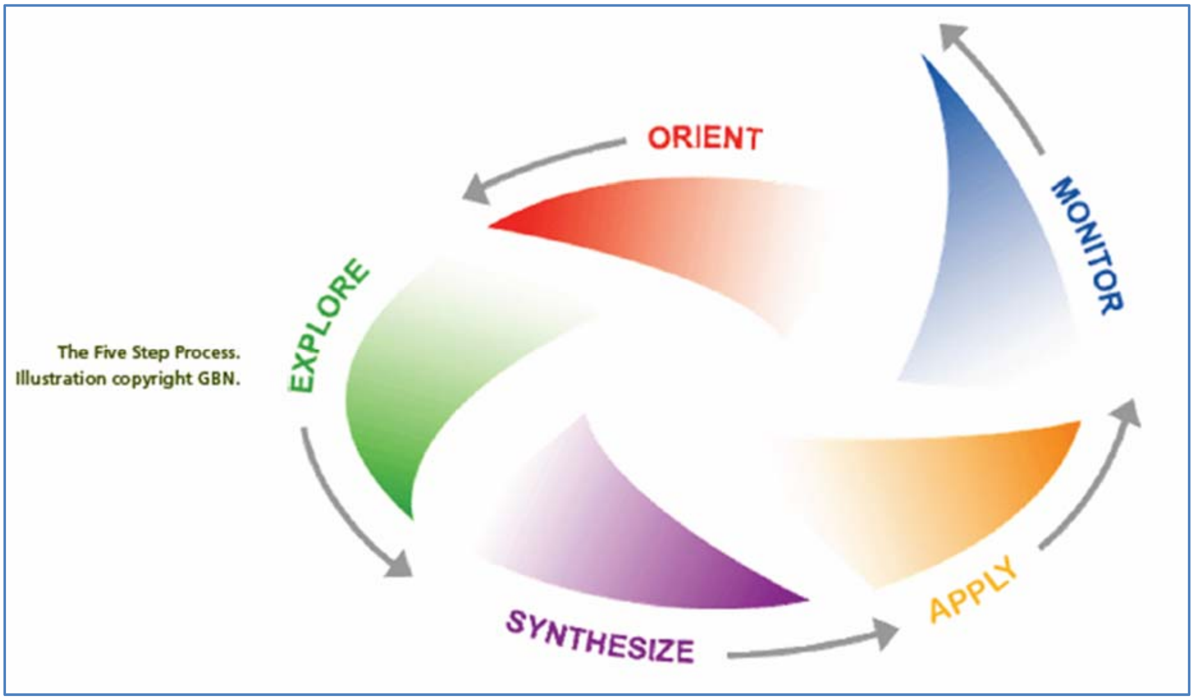

Figure 13: National Park Service five-step scenario planning process (*Using Scenarios to Explore Climate Change: A Handbook for Practitioners*, 2013)

5.0 Data-driven, science-informed solutions for resilience planning

5.1 Demise of climate stationarity

The demise of climate stationarity implies concomitant demise of statics solutions for climate resilience that are designed, implemented, and never re-evaluated again. Socio-environmental systems evolve over time and space**.** Natural and human "forcings" result in impacts to socio-environmental systems, which spur adaptation and mitigation responses, which in turn introduce feedbacks that interact with system stressors in non-linear and often unpredictable ways. This is depicted in a high-level conceptual framework used by the US Global Change Research Program (USGCRP) in its design of the National Climate Assessment national climate change indicators (Figure 14).

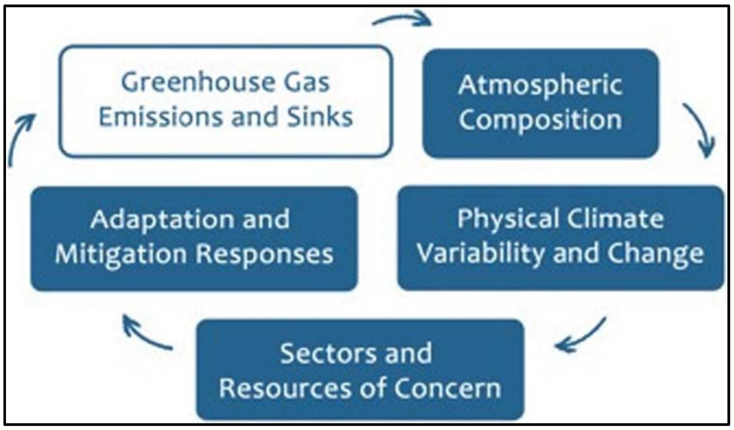

Figure 14: Conceptual model of the USGCRP indicator system (Kenney, Janetos, & Gerst, 2018)

The frameworks presented in Section 4.1 "Frameworks for understanding the socio-environmental decision landscape for resilience planning" are also indicative of the cyclic forcings-feedback mechanism inherent in socio-environmental systems. Readers may also be familiar with the Bretherton Diagram (Bretherton & others, 1988) that depicts a tightly coupled interconnected earth system where human and natural forcings and feedback interact in highly complex manners.

The evolution of socio-environmental systems in this manner necessitates the incorporation of a monitoring and evaluation (M&E) component when the construction and maintenance/operations phases of resilience projects have lifetimes that span decades. M&E facilitates adaptive management as the socio-environmental system evolves. Adaptive management is the new operational norm given the likely increasing variability observed in climatic variables.

5.2 Data-driven, science-informed, traceable solutions

5.2.1 Data to decisions pipeline

An earlier version of Figure 15 (DOI: 10.6084/m9.figshare.1287369.v4) was first presented before President Obama's Council of Advisors on Science and Technology (PCAST) in 2013 to demonstrate the role of informatics as the "pipeline" between data-intensive science and informed policies and decisions.

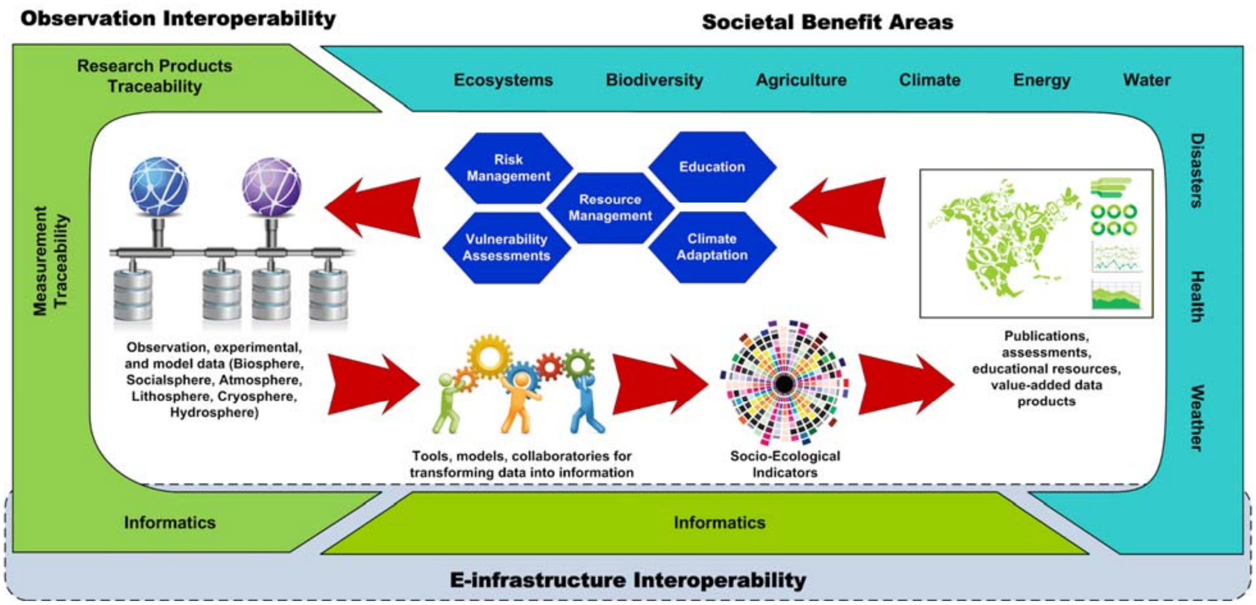

Figure 15: Schematic tracing the transformation of scientific data into decisions (Wee, 2018).

A Hudson River Project mitigation plan (corresponding to the blue hexagon in the figure above labeled "Climate Adaptation") that is **demonstrably traceable to data and science** would explicitly refer to model data that identifies flood-prone areas in Hoboken through the use of inundation models. There is an emphasis on "demonstrably traceable". Just as reproducibility in science can be achieved via a sufficiently detailed documentation of analysis protocols and the ability to inspect intermediate work products (including data, code, and documentation), traceable decisions should similarly be reproducible.

5.2.2 Decision opacity

Decisions (e.g. the decision behind the flood mitigation strategy selected for the Hudson River Project) are ultimately executed in a decisionmaker's head. It is arguably impossible in many cases to document the neural pathways by which the decisionmaker arrived at a decision. Similarly, deep learning techniques are not amendable for the derivation of an intuitive, human-understandable narrative about how and why the neural network outputs ended up that way (explanations of back-propagation schemes, the number of network layers, and neuron weight configurations are hardly intuitively appealing).

Reproducible science requires that research be described in sufficient detail for an independent party to replicate the work. Just as scientists may have reservations against full transparency with their methods in a scientific publication, decisionmakers are often not willing or able to document their decision processes for a myriad of reasons, including political considerations, threat of legal exposure on grounds of liability, the excessive effort required to maintain detailed documentation, etc.

Moreover, what constitutes a decision?

Section 5.2.3 describes a type of decision that, under United States law, requires parties to justify and describe the basis of environmentally sensitive decisions to the public. Section 5.2.4 describes a method from operational research and decision science that requires decisionmakers to explicitly model their decisions in a manner that is determined by data.

5.2.3 ROD: Decision documentation under the National Environmental Policy Act

A Record of Decision (ROD) is a document required under a number of federal regulations to fulfill specific requirements stipulated under the National Environmental Policy Act of 1969 (as amended). RODs are often made available for public comment, especially if such documents are drafted as part of an Environmental Impact Statement (EIS).

A ROD nominally includes a list of alternative actions that can be used to fulfill a set of stated objectives. For example, the objectives could be related to “decrease the impacts of flooding”, with each alternative describing different ways to implement measures that minimize flooding. The ROD needs to state which alternative was selected for implementation, with a justification for the decision. There is no known requirement to demonstrate traceability in ROD decisions to scientific analyses.

Figure 16 shows the contents page for the ROD for the Hudson River Project.

Record of Decision

Rebuild By Design: Resist, Delay, Store, Discharge Project
Cities of Hoboken, Weehawken, and Jersey City
Hudson County, New Jersey

TABLE OF CONTENTS

		Page
1.0	INTRODUCTION	1
2.0	PROJECT SUMMARY	3
3.0	DECISION	5
4.0	OTHER ALTERNATIVES CONSIDERED	9
5.0	ALTERNATIVES ANALYSIS	11
6.0	SUMMARY OF ENVIRONMENTAL IMPACTS	14

Figure 16: Hudson River Project Record of Decision

40 CFR § 1505.2 stipulates that the ROD shall (emphasis added):

> *(a) State what the **decision** was.*
>
> *(b) Identify all **alternatives** considered by the agency in reaching its decision, specifying the **alternative or alternatives which were considered** to be environmentally preferable. An agency may discuss preferences among alternatives based on **relevant factors including economic and technical considerations** and agency statutory missions. An agency shall identify and discuss all such factors including any essential considerations of national policy which were balanced by the agency in making its decision and state how those considerations entered into its decision.*
>
> *(c) State whether all practicable means to avoid or minimize environmental harm from the alternative selected have been adopted, and if not, why they were not. A monitoring and enforcement program shall be adopted and summarized where applicable for any mitigation.*

40 CFR § 1505.2(b) stipulates that any preference analysis may include "economic and technical considerations". Some of these "considerations" may actually have been borne through **technical analyses using observation and model data**, although there is no requirement for the ROD to provide a reference to those technical analyses.

RODs nevertheless may be used as the starting point to determine how and if decisions are supported by data-driven technical analyses. Moreover, the legal requirements related to RODs and the amount of resources invested in preparing such documents provide a rich source of documents that can be used to inform our technical requirements.

The EPA used to hold a contest (up till FY 2004) to identify well written RODs. These previously identified RODs could be used to further inform the d2dprovenance project of the types of decisions that would benefit from traceability.

Kornyshova & Deneckère, 2012 proposed a **Decision Making Ontology** that reflects the types of information that is required to be captured in a ROD. Their ontology is shown in Figure 17. Locher & Costa, 2017 propose an ontology called the **Multi-Entity Bayesian Decision Graph Decision Ontology**.

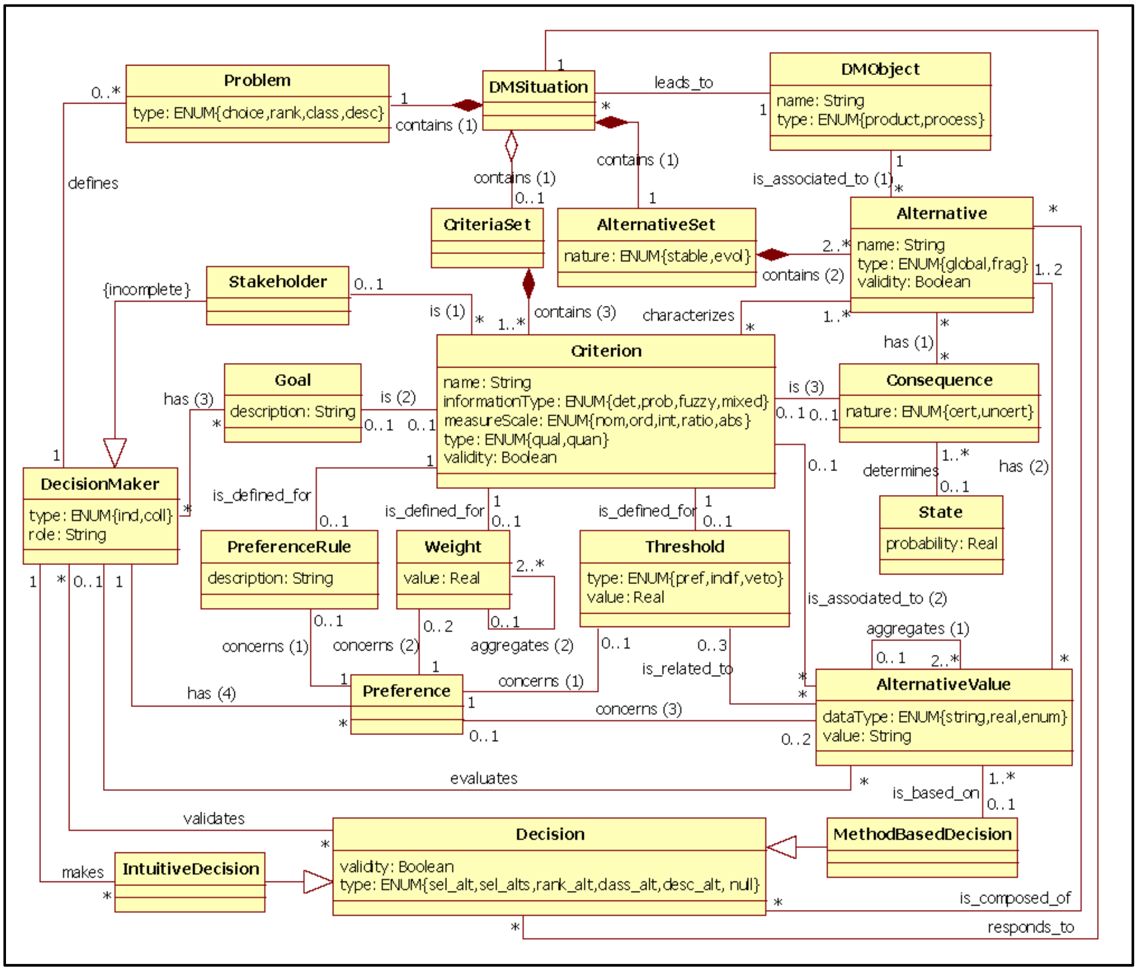

Figure 17: Decision Making Ontology (Kornyshova & Deneckère, 2012)

5.2.4 Incorporating data into decisions in a traceable manner using influence diagrams

Figure 18 shows an **influence diagram** (a tool often used in formal decision analysis like Structured Decision Making: see Section 5.3) for a hypothetical data-driven, science-informed coastal resilience project that incorporates green and grey infrastructure. The goal of modeling the decision as an influence diagram is to use model data to determine how stakeholder objectives (green nodes) can be accomplished via a slate of management options (gold nodes).

In this example, stakeholders have identified two priority objectives of extreme weather mitigation: to minimize damage to city infrastructure and to minimize salt-water intrusion into groundwater sources. There are various management options, that can be combined in varying degrees (corresponding to varying levels of capital investment), that can be combined to meet stakeholder objectives. The degree to which the stakeholder objectives are met is determined by how the management options are combined. A **decision model** incorporating an influence diagram like Figure 18 helps the decision maker arrive at a decision using model and observation data (pink nodes).

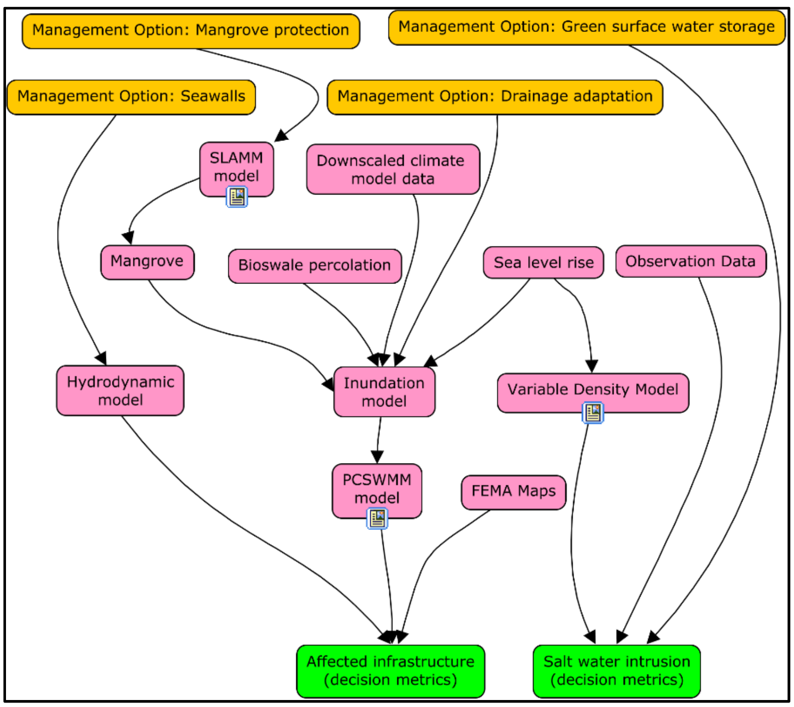

Figure 18: Hypothetical influence diagram for coastal resilience decision analysis.

5.2.5 Advantages of traceable climate resilience solutions

As outlined above, our understanding of socio-environmental systems will evolve as these socio-environmental impacts and stressors co-evolve. Decisions will likely need to be re-assessed in light of new data and science. When resilience building involves expensive, long-term capital investments in equipment and infrastructure, decisions almost invariably need to be revisited over time.

For example, the design of flood mitigation structures for an area within a city may need to be modified as municipal land-use code changes the distribution and percentage of impervious surface. That, combined with the demise of climate stationarity, increases the likelihood that there will be scenarios that have not been anticipated by planners when the plan was crafted.

Well-documented resilience plans that are traceable to data and science facilitates updates to mitigation plans using the most updated observation and model data that best characterizes the circumstances of that time.

Moreover, climate mitigation planning at the state and local levels will very likely continued to be conducted in an era of increased financial austerity. Traceable resilience solutions facilitate the sharing, reuse, and repurpose of resilience plans that are data-driven and science-informed. See also Section 5.3.2 “Integrating, managing, and disseminating decisions using SDM”.

5.2.6 State of resilience planning in the United States

At the time of writing, there appears to be extensive availability of web-accessible human-readable documents that correspond to the US CRT steps “Explore Climate Threats” and “Assess vulnerability and risks”. There is however a paucity of resources for activities related to “Prioritize Actions” and “Taking Action”. This pattern was also observed by Hansen et al., 2012 (Figure 19). This is perhaps not surprising, given that actual steps taken to implement plans are likely to be costly.

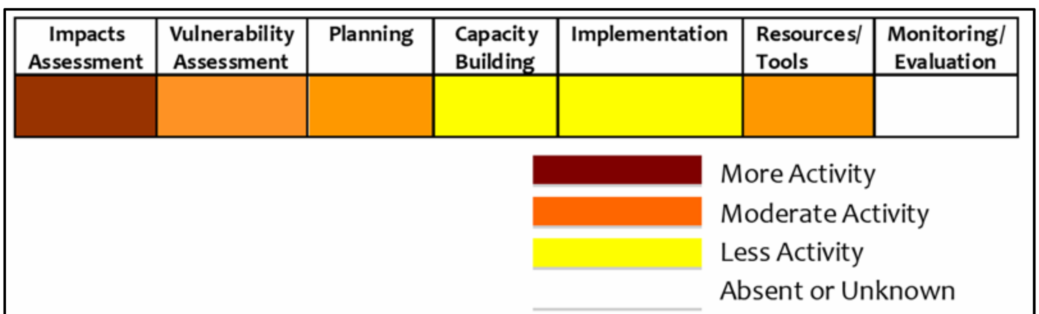

Figure 3. Activity levels vary in the stages of the adaptation process in the United States for all sectors. Significant effort has gone into problem identification (particularly impact assessments), some effort has gone into planning and tools development, there is insufficient effort toward capacity building and implementation, and there is essentially an absence of monitoring and evaluation of adaptation efficacy.

Figure 19: State of resilience planning in the US (Hansen et al., 2012)

This is not expected to change anytime soon. Fully traceable resilience plans may facilitate data-driven and science-informed planning, but will unlikely lessen the need for trained human capital to understand and assess the precedent steps in the planning process in order to repurpose an existing traceable resilience plan.

Hansen et al., 2012 point out that there is insufficient support for capacity building. Wee & Piña, 2019 made a similar point in the context of the “Resilience Genome Initiative” (see Section 6 “Long-term Vision”) by way of calling for the need to invest in “capacity building to foment a **transdisciplinary community of Resilience Genome Engineers**”. Piña & Wee, 2018 further maintain that:

data is necessary, but not sufficient, to empower a generation of problem solvers poised to confront socio-environmental challenges related to agroecosystem sustainability. What is needed, but has been largely absent, is a cadre of problem solvers who are transdisciplinary thinkers proficient in wrangling the multitude of resources available in the open digital commons.

5.3 Decision analysis for climate resilience using Structured Decision Making (SDM)

The concepts in the Kornyshova & Deneckère, 2012 ontology share similarities with a formal decision analysis method called Structured Decision Making (SDM) (Gregory et al., 2012). Figure 5 shows the relationship between SDM and the US CRT's "Steps to Resilience" planning framework. The SDM methodology is replicated below as Figure 20.

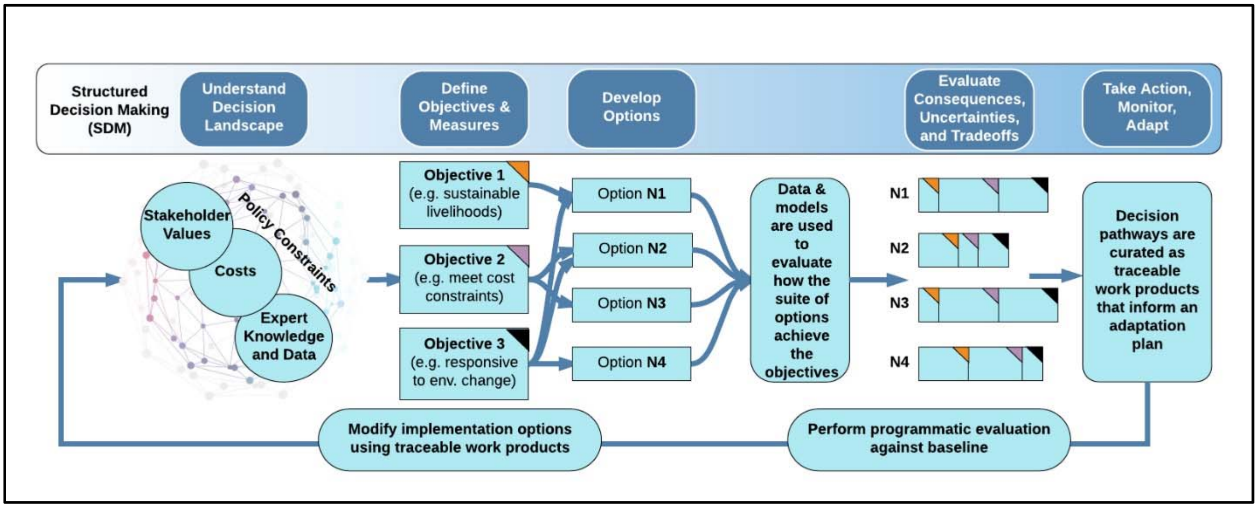

Figure 20: A schematic of Structured Decision Making.

5.3.1 Overview

SDM provides a transparent framework to develop solutions for climate resilience challenges. The method combines analytical methods from the decision sciences and applied ecology with insights drawn from cognitive psychology and the experience of facilitators and negotiators. Incorporating data and statistics using SDM begins with a comprehensive understanding of the decision landscape (e.g., programmatic objectives, desired outcomes, possible implementation options, regulatory aspects of the decision) (see also Section 4.1 "Frameworks for understanding the socio-environmental decision landscape for resilience planning").

Stakeholders and decision-makers are critical to identifying and defining the decision landscape, and are actively involved throughout the decision process. Stakeholders may also be seen as the custodians of values deemed paramount by the communities they represent. There is an emphasis on clearly communicating judgments about costs, values, uncertainty, and decision risks.

The values and preferences of stakeholders are made explicit and translated into decision objectives. Data and models are used (e.g. via an influence diagram in Figure 18) to quantify the effects of various implementation options. Probabilistic consequence modeling can then be used for forward and backward reasoning, sensitivity analyses used to identify influential variables, and value of information analyses used to address reducing sources of uncertainty.

5.3.2 Integrating, managing, and disseminating decisions using SDM

SDM was used for a “Coastal Community Resilience Planning and Decision Making” project in Dania Beach, Florida. The project included quantitative assessments of how different combinations of environmental management options impacted objectives derived from stakeholder values and cost constraints. That information was captured with a web-based SDM software that linked objectives, implementation options, performance measures, models, and data into a cohesive structure. These deliverables may be easily assembled into an electronic "decision-management package".

The increased emphasis on reproducibility, transparency, and traceability has changed the way science is done. Publications are increasingly released with accompanying data and code, often freely web accessible. The public would be well-served if the same paradigm is applied to climate resilience planning. If a resilience plan is released with an accompanying electronic decision-management package ("Package") that can be archived and distributed, the decision-provenance (a formalized, canonical representation of how decisions are made) embedded in a Package enables a number of capabilities summarized in Figure 21 (Wee et al., 2017).

A decision-management package can also be conceptualized as a “Resilience Genome” in the specific instance where suck a package is used for climate resilience applications.

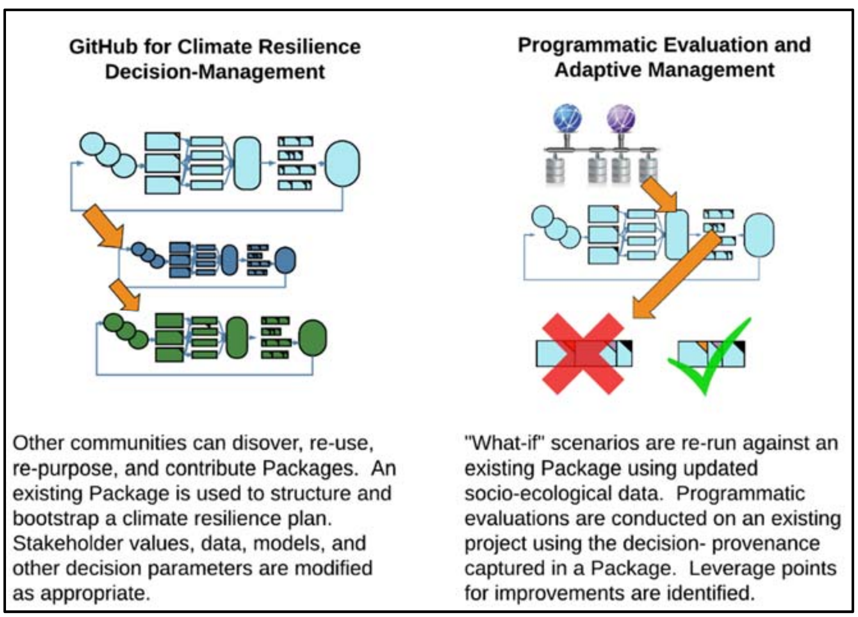

Figure 21: Decision Management Packages (Wee et al., 2017)

6.0 Long-term Vision

6.1 The Resilience Genome

Just as an organismal genome encodes most of the "building instructions" of an organism, a Resilience Genome (Wee & Piña, 2019) encodes the information required to reconstruct a climate resilience plan (Figure 22). A Resilience Genome contains the data-to-decisions provenance that enables the detailed planning steps to be recreated by connecting data, code, and information to decisions. Calls for increased reproducibility in science attest to the importance of provenance. A Resilience Genome accomplishes for decisionmakers what Jupyter Notebooks accomplishes for scientists.

The Hudson River Project Resilience Genome would thus represent an end-to-end trace of the decision process, abstracted to a sufficiently high level that allows discovery and comparison of other Resilience Genomes. A machine-readable version of the Hudson River Project Record of Decision (ROD) would include a Resilience Genome that encodes the decision components in the ROD, including:

- the resilience approach that was ultimately selected,
- the alternative approaches considered,
- the criteria for selecting from amongst the alternatives,
- "economic and technical considerations" (as per 40 CFR § 1505.2(b): see Section 5.2.3 "Decision documentation under the National Environmental Policy Act") through models and data.

Provenance must, however, be encoded in a machine-readable format to reap the benefits of machine-assisted search and reasoning. The US Global Change Research Program (USGCRP) demonstrated thought leadership by incorporating both human- and machine-readable provenance in its quadrennial national climate assessments (NCA) to establish traceability between an NCA scientific finding to supporting Digital Research Objects. Human-readable provenance is captured in the NCA's "traceable accounts", while machine-readable provenance is captured in the Global Change Information System (Ma, Fox, Tilmes, Jacobs, & Waple, 2014; Tilmes, 2012).

The Genome aims to extend that traceability into the vicinity of decisions that culminate from actions that are informed by scientific findings. By **aligning the Genome structure to resilience frameworks like the one used for the US Climate Resilience Toolkit** (Figure 22 and see also Section 4.2 "Frameworks for climate resilience planning"), we encode the entire data-to-decisions workflow: from the data used to assess climate threats and vulnerabilities, to the publications that analyze climate impacts, to stakeholder values, and to the decision-science models that integrate the previous information to produce a numerically ranked set of climate adaptation options.

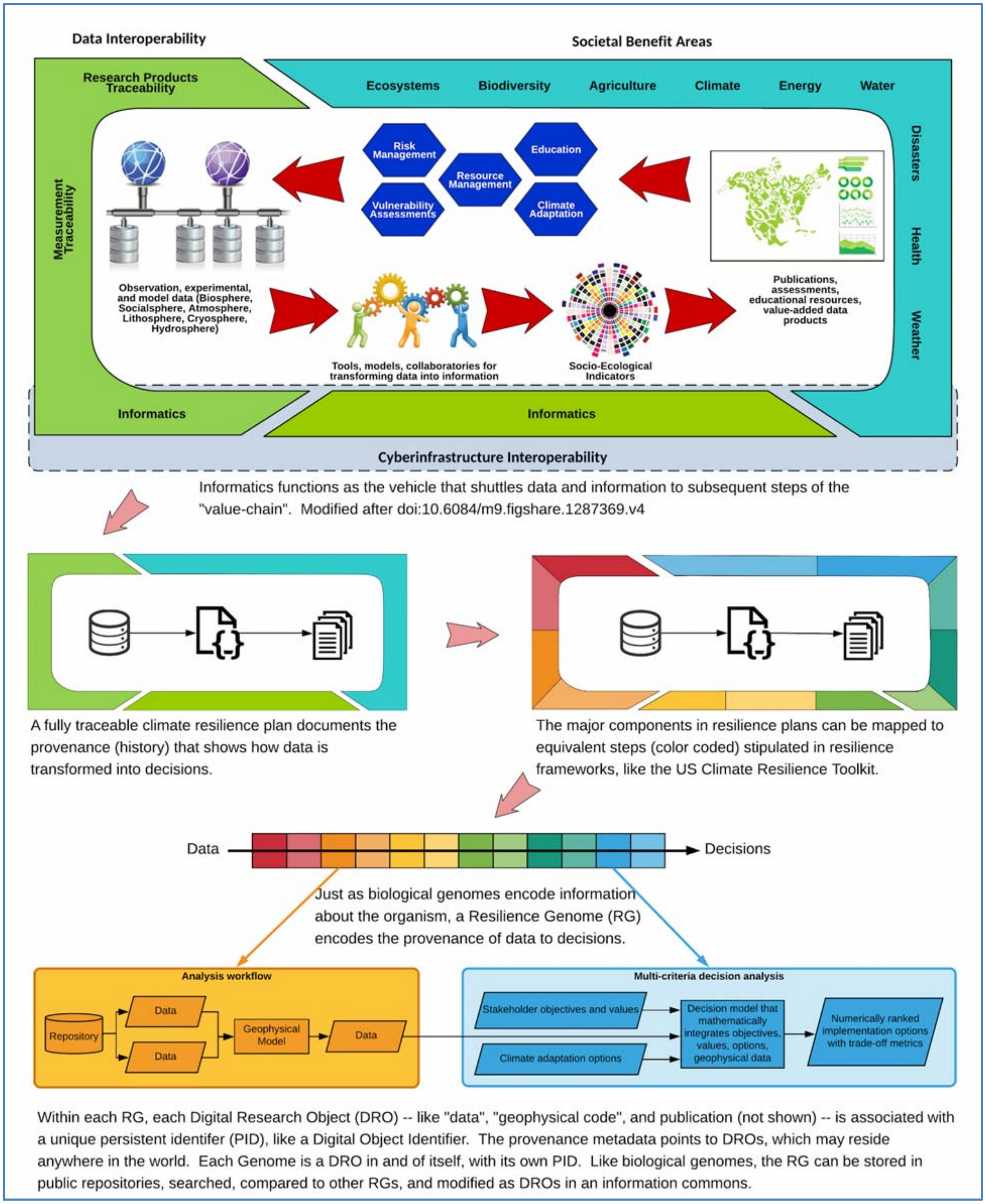

Figure 22: The Resilience Genome (Wee, 2019c).

6.2 Searching and using the Resilience Genome repository

How does the resilience community use a shared Resilience Genome library that curates traceable resilience plans? Bioinformaticians routinely run a target DNA sequence through a "DNA alignment tool" that searches a database for sequences that return a partial match. The same capability, augmented with semantic technologies, could be used to search a Genome repository (Figure 23).

Each Genome returned by the search engine comes replete with a full data-to-decisions provenance, including pointers to data, code, information, and decision models. Adapted Genomes could be contributed back to the repository and appropriately curated by the community using existing practices for managing open code. This process of selecting, modifying, and updating Resilience Genomes mirror "natural" evolutionary processes that lends an organic appeal to the way that the Genome is gradually shaped to meet societal needs.

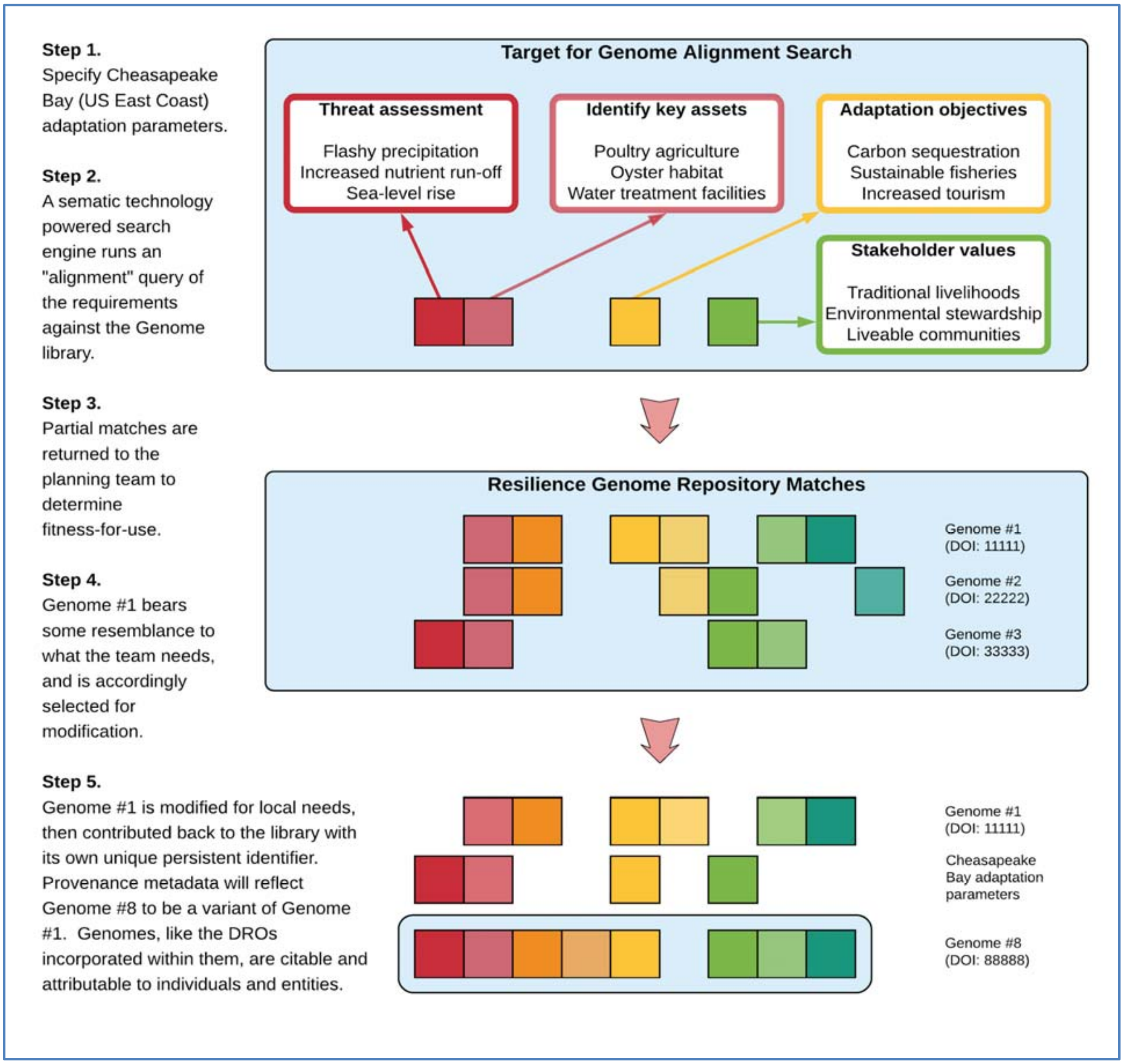

Figure 23: Searching and using the Resilience Genome Library (Wee, 2019e).

6.3 Knowledge graphs

6.3.1 Genomes as graphs

Resilience Genomes may be implemented as knowledge graphs (Figure 24).

Henceforth, the terms "Resilience Genomes" and "knowledge graphs" are used interchangeably. Any attempt at a technically defensible definition of the term "knowledge graph" within the scope of this report would not do it justice. A passable, but incomplete, definition of "knowledge graph" is a semantic network that represents highly specialized knowledge for a given knowledge domain structured using subject-predicate-object triples in a manner similar to that employed by Resource Description Framework (RDF) graphs.

Figure 25 and Figure 26 depict different views of how Resilience Genome Libraries that curate knowledge graphs can be populated and searched. Figure 25 and Figure 26 are variants of Figure 22 and Figure 23 respectively.

Figure 26 includes a "Genome Assembler". The Genome Assembler uses concepts extracted from documents (see Section 6.3.2 below) and assembles a data-to-decisions provenance trail implemented as a knowledge graph. The Genome Assembler will, initially, require humans to QA/QC automatically assembled, and probably incomplete, Genomes. As the corpus of Resilience Genomes grows over time, the Genome Assembler will more likely be capable of employing machine learning techniques to produce high-quality genome assemblies (i.e. knowledge graphs) that require minimal human edits.

Examples of knowledge graph implementations include the Google Knowledge Graph, Diffbot, and the graph-powered scientific literature tool iris.ai (Extance, 2018). These technologies are proprietary, are often designed for highly specific applications (e.g. modeling consumer behaviors using knowledge graphs that establish the relationships between people, products that they have purchased in the past, and the suppliers who provide those products). The proprietary and commercial nature of these applications may render the technology less accessible to the research community or to those who are exploring new ideas like those described in d2dprov. However, they are useful models to explore to consider how Resilience Genomes can be implemented as knowledge graphs.

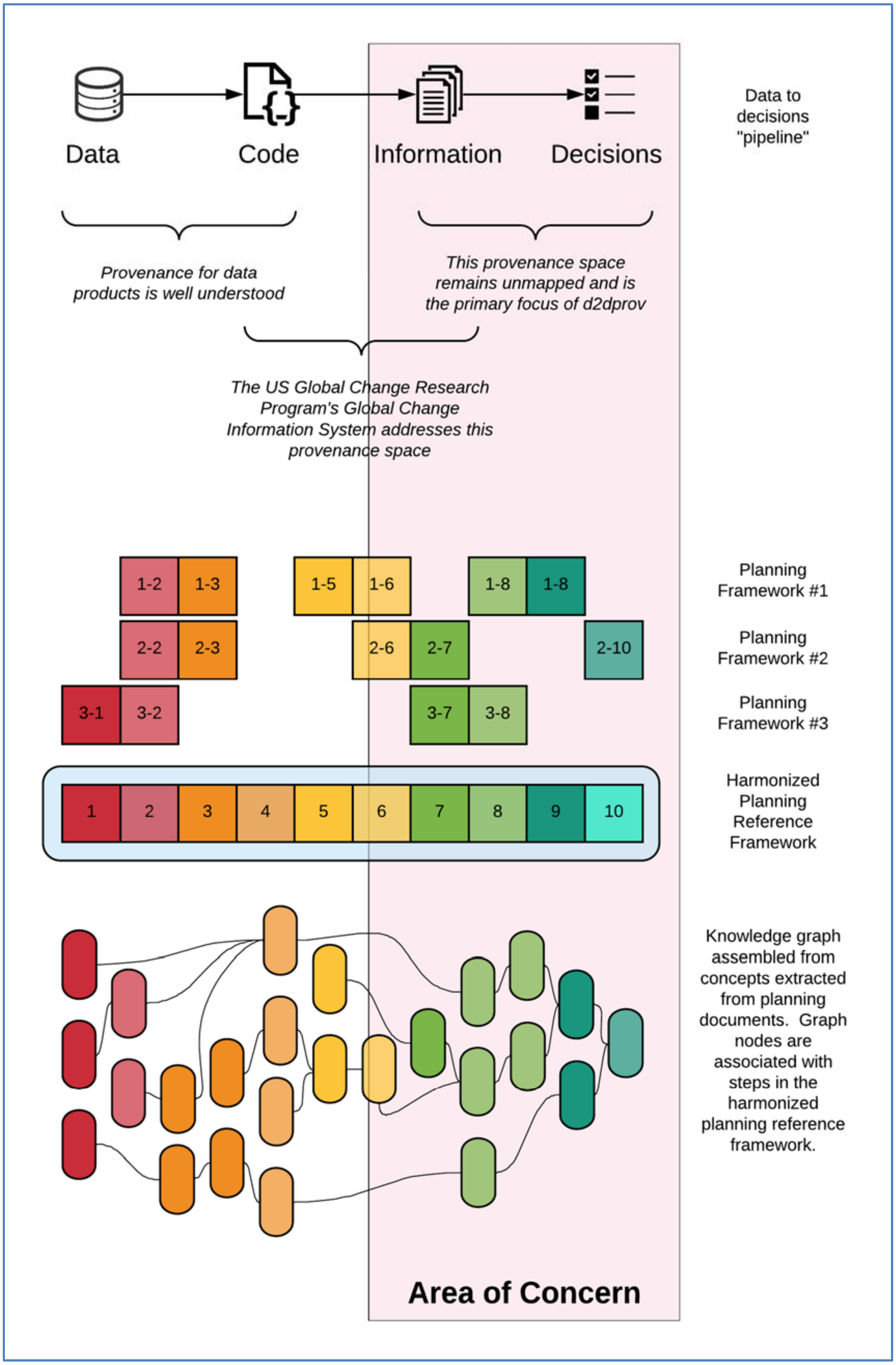

Figure 24: A Resilience Genome implemented as a knowledge graph.

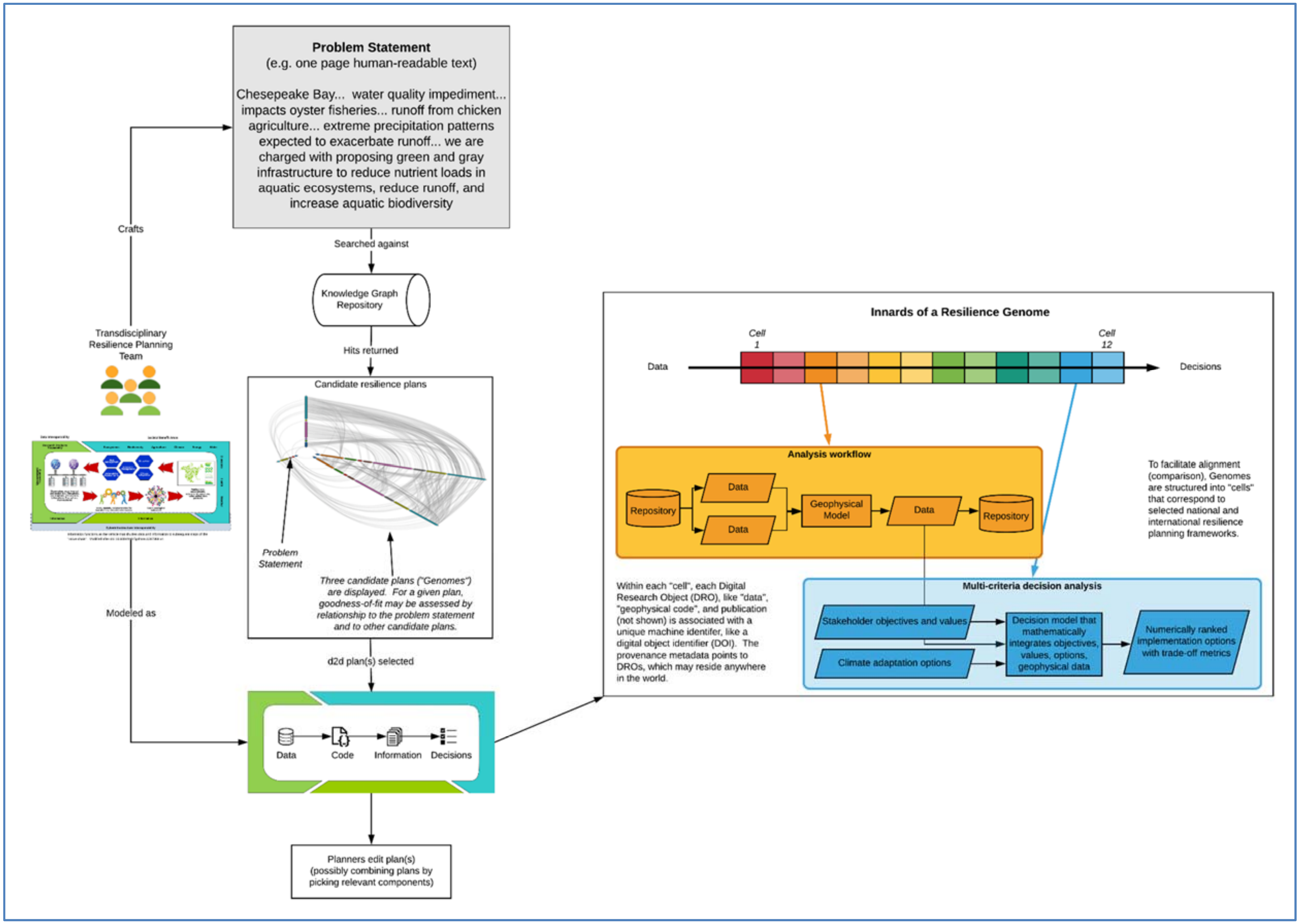

Figure 25: Planning team querying a Resilience Genome Library (Wee, 2019d)

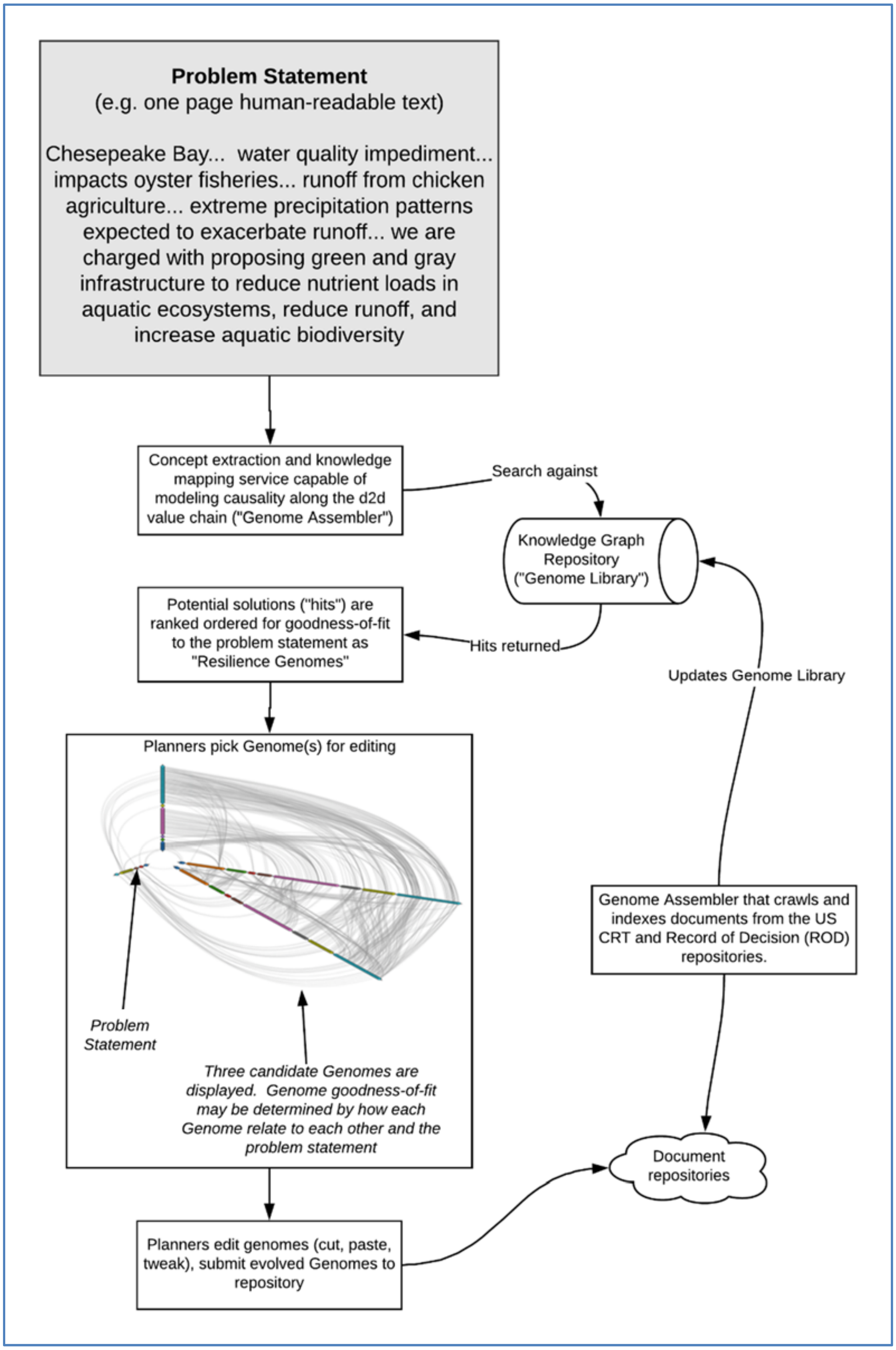

Figure 26: Resilience Genome Library query and update (Wee, 2019d)

6.3.2 d2dprov prototype concept extractor

D2dprov team member Tom Narock implemented a prototype concept extractor written in Python that parsed human-readable documents for terms that are defined in ontologies that are maintained by the informatics community (Narock et al., 2019). The following documents were text-mined:

1. a set of documents released in 2018 for a coastal and social resiliency initiative for the Tottenville Shoreline (Borough of Staten Island, Richmond County, New York), and

2. a set of documents released in 2017 for California's Rim Fire restoration initiative.

See Narock et al., 2019 and Wee, 2019a for more details.

6.3.3 Challenges in the automated generation of knowledge graphs

The provenance space between "data" and "information" in Figure 24 is relatively well understood. Knowledge graph generation for that space is enabled by existing ontologies that model how data, scientific models, and information are related.

Fully automated generation of provenance to link data, models, and information will however remain problematic for the foreseeable future. For example, the US Global Change Information System (GCIS) team has had little choice but to manually disambiguate the identities of journal publication authors in order to establish traceability between assertions in the National Climate Assessment and the scientific publications that support those assertions (personal conversations with GCIS team).

Although it is resource-intensive to manually edit and curate the provenance space between "data" and "information" because full automation will likely remain unattainable for some time, provenance generation is a tractable problem. **By comparison, the automated assembly of extracted concepts into knowledge graphs for the space labeled "Area of Concern" in Figure 24 that connects "information" to "decisions" represents a severe challenge.**

The challenge may be stated as:

> Given a collection of human-readable documents with no a-priori assumption of how those documents are organized, and where no documents are coupled with semantic annotations, how can one implement an automated process that assembles extracted concepts into a traceable account that leads from decisions back to information?

The above challenge may be further elaborated as follows:

1. **No assumptions can be made about how information is structured within a document.** Although one may assume that a given document is structured into sections, no assumption can be made about delineation between those sections. For example, are sections indicated by a numbered sequence (as in this document), or are sections indicated by text that is visually formatted differently without section numbers, or are sections separated by some type of delineator like a series of repeating delimiter characters (e.g. "------", "…….", "++++++", "=======", "#######")?

2. **No assumptions can be made about how information is structured between documents.** Even if we were to assume that a planning project were staffed with a dedicated human document manager (e.g. to oversee the assignment of document IDs, file naming conventions, etc), a dedicated human QA/QC manager (e.g. to perform conformity checks on documents to see that they adhere to project standards), and document management software (e.g. GIT, CVS, or SVN), no assumptions can be made about how the documents relate to one another in the context of a planning framework.

3. **Level of granularity of representing the processes that led from "information" to "decisions".** The accurate automated extraction and representation of fine-scale decision processes that simultaneously may be (a) embedded in various places within a document and also (b) distributed across several documents will be challenging. The automated extraction and representation of coarse, high-level decision processes may be somewhat easier.

6.3.4 Ways to ameliorate challenges in the "Area of Concern"

Means to ameliorate these challenges include:

1. **Use a decision-making ontology to inform options for semantic annotation.** There exists a number of decision-making ontologies and decision-making schemas (schemas that represent information models that have yet to be implemented as a formal ontology). A brief review of selected ontologies / schemas is provided in Narock et al., 2019 and Wee, 2019a.

   These ontologies / schemas can be used by an automated Genome Assembler to guide its assembly of extracted concepts into a knowledge graph. A decision-making schema like that proposed by Kornyshova & Deneckère, 2012 models the decision-making process at a very fine-grained scale, in contrast to an ontology like that proposed by Car, 2017 that focuses on decision-making process at a higher conceptual level.

   One might posit that a light-weight ontology along the lines of Car, 2017 would perform better when employed by a fully automated Genome Assembler, because modeling processes at a higher-level allows for more wiggle room for the algorithm to get things "right".

   On the other hand, the fine-grained Kornyshova & Deneckère, 2012 schema would be more useful for a Genome Assembler that involves human judgements to tweak and curate the data-to-decisions provenance.

   This is a technical means to ameliorate the challenges of Genome Assembly for the "Area of Concern". The following solutions employ a combination of technology tools and policy imperatives.

2. **Human-mediated semantic annotation.** Decision-making ontologies could be used to perform an initial parse of a set of planning documents. The resulting knowledge graph can then be made available for human inspection and tweaking.

Alternatively, document authors could be incentivized to semantically annotate their human-readable documents, in a fashion similar to populating metadata fields when one deposits a dataset into a repository. But, would the benefits of increased discovery and citation be sufficient to incentivize professionals to semantically annotate their documents?

Another option would be to make semantic annotation of documents mandatory for grant recipients, regardless of whether it is a grant for intramural research within a government agency, or extramural research conducted by external parties.

One could mandate that authors semantically annotate documents as a prerequisite to the publication, promulgation, or archival of documents.

Any of these options would likely elicit strong reactions from the community. Implementing any means of human-mediated semantic annotation will almost certainly impose additional administrative burdens. Individuals will have to learn to use annotation tools and budget for additional effort to annotate documents. The impression of administrative burden may however be somewhat ameliorated by distributing semantic annotation tasks along the entire resilience planning process.

3. **Use standardized resilience planning vocabulary.** Federal agencies that disburse grants for mitigation planning should also encourage, or require, grant recipients to produce publicly accessible documents that are structured into sections using vocabulary that reflect that agency's preferred mitigation planning framework. This type of standardization will help improve the accuracy of Genome Assembly. See Wee, 2019a for details.

   This proposed solution, like the one before (solution #2 above), incorporates policy imperatives that will elicit reactions from the community.

7.0 Bibliography

Bloomberg, M. (2013). *Coastal Climate Resilience: Urban Waterfront Adaptive Strategies*. New York, NY.

Bretherton, F., & others. (1988). Earth System Sciences: A Closer View. *Earth System Sciences Committee, NASA*.

Car, N. J. (2017). *Modelling causes for actions with the Decision and PROV ontologies*. Retrieved from http://promsns.org/def/decprov

Climate-resilient development: A framework for understanding and addressing climate change. (2014). Retrieved from https://pdf.usaid.gov/pdf_docs/PBAAA245.pdf

Extance, A. (2018). How AI technology can tame the scientific literature. *Nature*, *561*(7722), 273–274. https://doi.org/10.1038/d41586-018-06617-5

Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., & Ohlson, D. (2012). *Structured Decision Making: A Practical Guide to Environmental Management Choices*. John Wiley & Sons.

Hansen, L., Gregg, R. M., Arroyo, V., Ellsworth, S., Jackson, L., & Snover, A. (2012). *The State of Adaptation in the United States*. Retrieved from https://www.macfound.org/media/article_pdfs/The_State_of_Adaptation_in_the_United_States.pdf

Kenney, M. A., Janetos, A. C., & Gerst, M. D. (2018). A framework for national climate indicators. *Climatic Change*. https://doi.org/10.1007/s10584-018-2307-y

Kornyshova, E., & Deneckère, R. (2012). Using an ontology for modeling decision-making knowledge. *Frontiers in Artificial Intelligence and Applications*, *243*(September 2012), 1553–1562. https://doi.org/10.3233/978-1-61499-105-2-1553

Locher, M., & Costa, P. C. G. (2017). The multi-entity decision graph decision ontology: A decision ontology for fusion support. In *Information Fusion (Fusion), 2017 20th International Conference on* (pp. 1–8).

Ma, X., Fox, P., Tilmes, C., Jacobs, K., & Waple, A. (2014). Capturing provenance of global change information. *Nature Climate Change*, *4*(6), 409–413. https://doi.org/10.1038/nclimate2141

Mueller, M., Spangler, T., & Alexander, S. (2013). Mid Semester Break.pdf. Retrieved from https://www.agrilinks.org/sites/default/files/resource/files/FTF Learning_Agenda_Community_Resilience_Oct 2013.pdf

Narock, T., Wee, B., Hoebelheinrich, N., Albayrak, R., & Teng, B. (2019, February 20). ESIPFed/d2dprovenance: Data to Decisions Provenance. https://doi.org/10.5281/ZENODO.2574103

New Jersey Department of Environmental Protection, N. J. (2017). *Rebuild By Design – Hudson River Final Environmental Impact Statement - Executive Summary*. Retrieved from https://www.nj.gov/dep/floodresilience/docs/rbdh-feis/executive-summary-rbd-hr-feis.pdf

Piña, A., & Wee, B. (2018). Transdisciplinary thinking essential to enable sustainable data-intensive agroecosystems. Retrieved from https://bigdata.cgiar.org/transdisciplinary-thinking-essential-to-enable-sustainable-data-intensive-agroecosystems/

Tilmes, C. (2012). Provenance Representation in the Global Change Information System (GCIS) (pp. 246–248). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34222-6_28

UKCIP risk framework | UKCIP. (n.d.). Retrieved January 18, 2019, from https://www.ukcip.org.uk/wizard/about-the-wizard/ukcip-risk-framework/

Using Scenarios to Explore Climate Change: A Handbook for Practitioners. (2013). Retrieved from https://www.nps.gov/subjects/climatechange/upload/CCScenariosHandbookJuly2013.pdf

Wee, B. (2018). Transformation of Data for Societal Benefits. https://doi.org/10.6084/m9.figshare.1287369.v4

Wee, B. (2019a). *D2dprov: Statement of Needs 2022. Technology and policy requirements to fulfill Vision 2025's proposed approach to data-driven, science-informed climate resilience decisions.* Washington, DC. https://doi.org/10.6084/m9.figshare.7692038

Wee, B. (2019b). *D2dprov: Vision 2025. A transdisciplinary science, technology, and policy vision for data-driven, science-informed resilience planning for 2025 and beyond.* Washington, DC. https://doi.org/10.6084/m9.figshare.7591238

Wee, B. (2019c). Encoding provenance of data to decisions using a Resilience Genome. https://doi.org/10.6084/m9.figshare.7594334.v1

Wee, B. (2019d). Resilience Genome: Vision for Creation, Use, and Evolution. https://doi.org/10.6084/m9.figshare.7605566.v1

Wee, B. (2019e). Resilience Genome Use Case. https://doi.org/10.6084/m9.figshare.7594355.v1

Wee, B. (2019f). US Climate Resilience Toolkit planning protocol mapped to Structured Decision Making. https://doi.org/10.6084/m9.figshare.7605632.v1

Wee, B., Black, P., Billig, P., Black, K., Duffy, P., Rupp, S., & Stockton, T. (2017). Data-Driven Decision-Management: A Values-focused Approach to Enable Traceable Decision Analytics for Adaptive Climate Resilience. https://doi.org/10.6084/m9.figshare.4515722.v2

Wee, B., & Piña, A. (2019). (IN PRESS) The Resilience Genome Initiative: A vision for adapting at the pace of socio-environmental change. *Eos, Transactions American Geophysical Union*.