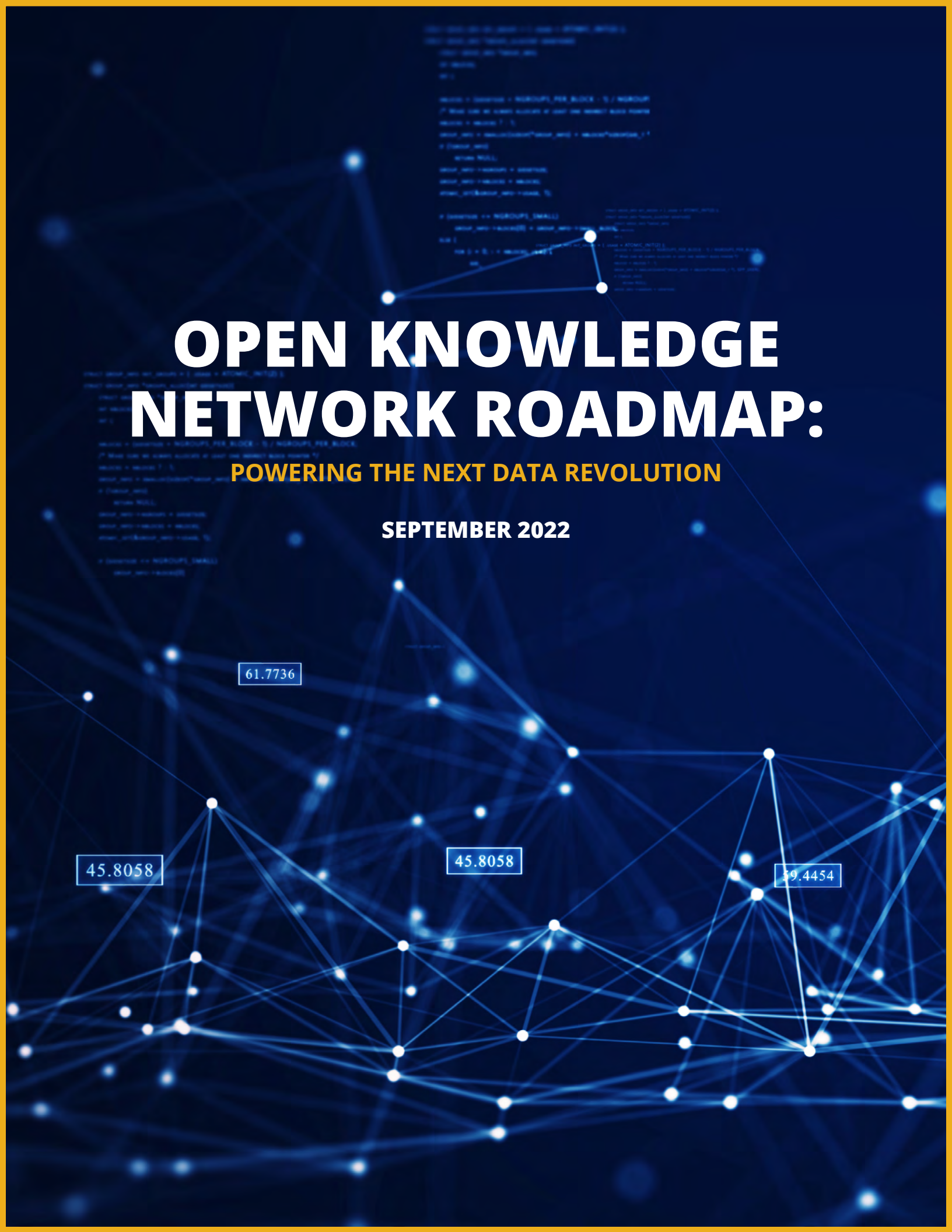


OPEN KNOWLEDGE NETWORK ROADMAP:

POWERING THE NEXT DATA REVOLUTION

SEPTEMBER 2022

AUTHORS

OKN Innovation Sprint Organizing Committee

Chaitan Baru
University of California, San Diego

Lara Campbell
National Science Foundation

Wo Chang
National Institute of Standards and Technology

Tess DeBlanc-Knowles
Office of Science and Technology Policy / National Science Foundation

Jemin George
National Science Foundation / DEVCOM Army Research Laboratory

Martin Halbert
National Science Foundation

Co-Authors

Kat Albrecht
Georgia State University

Luis Amaral
Northwestern University

Nariman Ammar
University of Tennessee

Todd Bacastow
Maxar Technologies

Sergio Baranzini
University of California, San Francisco

Matt Bishop
Open City Labs

Michael Cafarella
Massachusetts Institute of Technology

Silviu Cucerzan
Microsoft Corporation

Ying Ding
University of Texas at Austin

Brian Handspicker
Open City Labs

Oktie Hassanzadeh
IBM Research

Pascal Hitzler
Kansas State University

Florence Hudson
Columbia University

Sharat Israni
University of California, San Francisco

Angela Rizk-Jackson
University of California, San Francisco

Esther Jackson
Columbia University

Eric Jahn
Alexandria Consulting

Krzystof Janowicz
University of California, Santa Barbara

Bandana Kar
Oak Ridge National Laboratory

Sam Klein
Massachusetts Institute of Technology

Matthew Lange
International Center for Food Ontology Operability Data and Semantics

Ora Lassila
Amazon

Chengkai Li
University of Texas at Arlington

Ryan McGranaghan
Atmospheric and Space Technology Research Associates

Murat Omay
U.S. Department of Transportation

Adam Pah
Northwestern University

Co-Authors

Louiqa Raschid
University of Maryland

Greg Seaton
SierraLogic LLC

Paul Wormeli
Wormeli Consulting LLC

Glenn Ricart
US Ignite

Cogan Shimizu
Kansas State University

Lilit Yeghiazarian
University of Cincinnati

Emanuel Sallinger
Vienna University of Technology / Oxford University

Amanda Stathopoulos
Northwestern University

Ellie Young
Common Action

Technical Writers/Editors

Pamela Livingston
Writing Assistance Inc.

Douglas Maughan
National Science Foundation

Shelby Smith
National Science Foundation

ACKNOWLEDGEMENT

This document attempts to capture the insights of a large and diverse set of stakeholders who shared their time and insights over many months.

The authors are grateful for the enthusiasm and compelling expertise of the participants in the Open Knolwedge Network Innovation Sprint, who from February to June of 2022 attended six workshops and numerous smaller meetings. Many thanks are due to the federal agencies, companies, universities, non-profits, and other organizations who allowed their staff to join this effort. The authors particularly appreciate the enabling work of the NSF Track A: Open Knowledge Network grantees whose work has shown the power and promise of an OKN as well as the need to build a larger community.

The authors also gratefully acknowledge the many prior efforts that are referenced in this document and those that are not. We hope that participants from those prior efforts, participants from the OKN Innovation Sprint, everyone who reads this report, and other interested parties will continue to help advance the effort to create an Open Knowledge Network.


36.7702
30.7317
29.4454

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Open access to shared information is essential for the development and evolution of artificial intelligence (AI) and AI-powered solutions needed to address the complex challenges facing the nation and the world. The Open Knowledge Network (OKN), an interconnected network of knowledge graphs, would provide an essential public-data infrastructure for enabling an AI-driven future. It would facilitate the integration of diverse data needed to develop solutions to drive continued strong economic growth, expand opportunities, and address complex problems from climate change to social equity. The OKN Roadmap describes the key characteristics of the OKN and essential considerations in taking the effort forward in an effective and sustainable manner.

Data, information, and knowledge

Harnessing the vast amounts of data generated in every sphere of life and transforming them into useful, actionable information and knowledge is crucial to the efficient functioning of a modern society. Data and information should be easy to find, access, and reuse. Knowledge structures that enable integration of vast amounts of diverse data in service of a very broad range of uses across society represent the national infrastructure for a prosperous future. *Knowledge graphs* are an important type of knowledge structure to enable such integration. They consist of *nodes* and *edges* — where *nodes* represent real-world entities (e.g., a city, a neighborhood, a court case, a gene, a chemical compound), and edges represent different types of relationships among nodes. The Open Knowledge Network is envisioned as an open, interconnected network of knowledge graphs that serves as public, accessible infrastructure. This infrastructure will enable development of a variety of solutions for a broad spectrum of societal use cases using open, public data, as well as data requiring controlled access.

In February 2022, the National Science Foundation (NSF) in partnership with the White House Office of Science and Technology Policy, launched an *OKN Innovation Sprint*. The Sprint harnessed the collective insights of roughly 150 experts from government, industry, academia, and nonprofit organizations to help build a roadmap for a Prototype OKN (*Proto-OKN*), from specific use cases and end-user perspectives. The main findings from this Sprint are summarized in this report. The context for this activity is provided by the original vision for an OKN described in the NSF Harnessing the Data Revolution Big Idea and by projects in Track A: Open Knowledge Network of the **NSF Convergence Accelerator**.

Discussions during the Innovation Sprint helped clarify that the creation of an OKN is fundamentally a sociotechnical effort that must consider human, social and organizational factors, as well as a technical effort. Deep engagement is necessary among domain knowledge experts and a host of other stakeholders including data owners, decision-makers, various end-user communities, tool builders, and

knowledge representation experts. User-centered design approaches, stakeholder involvement and alignment, and customer engagement are essential to ensuring an impactful and sustainable outcome. The Innovation Sprint focused on the needs and requirements of end-users and various stakeholders. Sprint participants formed 17 use-case groups. Thirteen groups focused on "verticals" (i.e., specific application areas), and four focused on cross-cutting, "horizontal" themes. **Section 6** of the report introduces these use cases. Appendix A describes their activities and outcomes in detail.

Report outline

The OKN Roadmap report first introduces the OKN concept (**Section 1**), and prior initiatives related to this effort (**Section 2**). The essential characteristics of the OKN include an open system architecture, a dynamic system reflecting real-world information updates, the ability to link diverse information and deduce linkages among related information elements, and a simultaneous focus on the use case-driven "vertical" aspects as well as the technology-driven "horizontal" aspects. Moreover, critical aspects of the system include a governance structure to allow OKN stakeholders to direct and oversee policies, processes to ensure ethical use, methods to track data provenance, flexibility to scale up to support additional data and users, interoperability with new data sources and systems, sustainability to meet current and future end-user needs, enforcement of data access rights, and procedures and metrics for data validation (**Section 3**).

The activities needed to create the OKN include gathering requirements from various use cases, establishing the necessary information structures such as ontologies and schemas, working with existing repositories to link them to the OKN, fostering development of information structures to facilitate interconnecting data across domains, developing working prototypes of the system for specific use cases, facilitating engagement by subject matter experts, developing a variety of user-friendly interfaces for a broad range of users, and ensuring that it is possible to incorporate public as well as sensitive, access-controlled data. Data and Information flow within the OKN could be considered in several distinct phases which include *design* of appropriate information structures for proper data representation, data *ingestion* into the system, data *enrichment* to extract information from data, data *storage*, data use/*consumption* by various applications defined by different use cases, and on-going maintenance of the whole enterprise (**Section 4**).

In moving forward, the efforts to construct an OKN must be guided by the sociotechnical nature of the undertaking. Strong stakeholder engagement essential for the success of this effort can be built through practices like the participatory design approach that engages stakeholders as well as end-users and developing educational and training materials suited to the broad range of stakeholders. Stakeholder engagements can be further strengthened by enabling open contributions for a range of sources to promote transparency and employing a community-centric approach to ensure an ethical and responsible system. Long-term success of the effort would require building a committed community of stakeholders to secure sustainability, ensuring that new data and end-users are able to connect to and use the OKN, and building the connective fabric of OKN to prevent future fragmentation of the system (**Section 5**).

The Sprint participants recognized that the construction and deployment of OKN would be a community-based effort to develop a national-scale data infrastructure with the costs distributed across many stakeholders. The schedule for a Proto-OKN would take into account the possibility of leveraging the available technology and experience base, while also paying attention to the significant new needs and requirements uncovered during the Sprint process.

Looking ahead

The OKN is an essential element of a national AI infrastructure. It would provide users across all sectors — including government agencies, private organizations, academia, and the public — access to integrated information for a variety of uses, including tackling societal problems, driving evidence-based policies, and developing novel AI capabilities.

This is the opportune time to embark upon development of the Proto-OKN by embracing state-of-the-art technologies, leveraging related efforts, and building upon the many partnerships that have developed via the NSF Convergence Accelerator and over the course of the OKN Innovation Sprint.

OPEN KNOWLEDGE NETWORK ROADMAP:

POWERING THE NEXT DATA REVOLUTION

The Open Knowledge Network (OKN) Roadmap paves the way for the rapid deployment of an Open Knowledge Network — an essential infrastructure for enabling an artificial intelligence (AI)-driven future. The open infrastructure would enable sharing of open, public data, and potentially data requiring controlled access. It would provide the knowledge infrastructure necessary for integrating the diverse data needed to continue strong economic growth, expand opportunities, and address complex problems from climate change to misinformation, disruptions from pandemics, and advancing economic equity and diversity. Access to rich, structured data is essential for evolution and use of AI and AI-based methods and solutions to the complex challenges facing society today. The OKN Roadmap describes the key characteristics of the OKN and essential considerations in taking the effort forward in an effective and sustainable manner.

In 2017, the National Science Foundation (NSF), in its **Harnessing the Data Revolution Big Idea**, recognized the need for an open networked structure for linking disparate, heterogeneous information from diverse sources to yield novel data-driven insights and solutions. As a first step, in Fall 2019, NSF funded 21 OKN projects as part of the **Convergence Accelerator Track A** program. In Fall 2020, five of these projects entered a 2-year Phase 2 effort. In February 2022, the NSF and the **Office of Science and Technology Policy (OSTP)** engaged a larger community of subject matter experts, end-users, and stakeholders from government, industry, academia, nonprofits and other communities in a 4-month Open Knowledge Network Innovation Sprint. This activity culminated in a Proto-OKN Roadmap workshop in June 2022, leading to this OKN Roadmap report.

This OKN Roadmap report guides readers through the various considerations in creating and deploying an OKN. It provides background information demonstrating the compelling need for an OKN. It specifies needs and requirements to serve various stakeholder interests, ranging from climate change to the criminal justice system, developed through design exercises focused on a diverse set of use cases. The report is comprised of the following eight sections:

- **Section 1** introduces the vision of the OKN, including its features, functions, and benefits, and related work in this area.
- **Section 2** describes the initiation of the OKN activity and the OKN Innovation Sprint process.
- **Section 3** describes the key takeaways related to the characteristics of the OKN.
- **Section 4** describes the range of issues to be considered in creating an OKN.
- **Section 5** describes considerations for taking this effort forward in an effective and sustainable way.
- **Section 6** provides an overview of 17 use cases developed during the Innovation Sprint.

- **Section 7** describes a possible timeline for implementing a Proto-OKN.
- **Section 8** provides a conclusion based on the findings of the previous seven sections.
- Appendix A provides a detailed account on all 17 use cases from the OKN Innovation Sprint.
- Appendix B provides an overview of the five OKN Phase 2 projects supported by the NSF Convergence Accelerator's **Track A Program**.

1. Envisioning the OKN

The vision for the OKN is the creation of a common infrastructure that is driven by specific use cases but takes the form of a shared, general platform. The OKN is envisioned to transform our ability to unlock actionable insights from data, by semantically linking information about related entities.

The Internet – which began as an attempt to link files and then evolved into a digital infrastructure that serves as the backbone for modern life – and the related idea of *Semantic Web* (Berners-Lee et al., 2001), can be seen as an inspiration and model for OKN development. The OKN is envisioned as an ethical, trustworthy network of *interconnected knowledge graphs* (Chaudhari et al., 2022). It would provide a trusted platform for users accessing information as well as for users who are information providers by establishing and following ethical standards for data exchange and analysis.

Knowledge graphs — founded on the principle of applying a graph-based abstraction to data — have emerged as a compelling concept for integrating information and extracting value from multiple diverse sources of data at large scale (Hogan et al., 2022). Some of the largest knowledge graphs in existence are powering consumer applications including web search, e-commerce, advertising placement, and question-answering (**Noy et al., 2019**). These same technologies can be used to create an open platform, with open as well as access-controlled data, to develop impactful new applications for evidence-based policymaking, game-changing research, and many other key areas of societal impact.

The benefits include the ability to provide answers to questions that might otherwise require inordinate effort in assembling, integrating, and analyzing datasets curated by different organizations. For example, questions such as those listed below could be answered quickly by the OKN.

- "Have there been unusual clusters of earthquakes in the United States in the past six months?"
- "What is the best combination of chemotherapeutic drugs for a 56-year-old female with stage 3 brain cancer?"

Providing a platform for shared access and cross-sector information-sharing

Envisioned as an inclusive, open, community-driven infrastructure accessible to all, the OKN would not only provide a trusted platform to empower a host of new applications, but it would also open new vistas in AI and data science research, including in fairness, bias, diversity, equity and inclusion. Access to OKN infrastructure and content would allow researchers and practitioners to develop more robust and efficient approaches to answering questions, more expressive frameworks to capture knowledge, and more natural interfaces to access that knowledge (**NSCAI Final Report, 2021**). Any organization

regardless of size or sector could benefit via a multi-sector, community-based effort that would help share the burden of development via open-source software and open, shared standards.

Considering factors and approaches for OKN design

Stand up of an OKN would leverage prior work in **Semantic Web** and **Linked Open Data**, and make full use of the experience and the platform provided by the highly successful **Wikidata** effort, as well as other more recent data commons efforts.

Many of the core technologies that underlay the OKN are well-established (Chaudhari et al. 2022; Hitzler 2021; Hogan, et al. 2022; Gutierrez, 2021). However, successful creation of the OKN is much more a sociotechnical challenge (Baxter and Sommerville, 2011) than merely a technical exercise. The design and implementation of an OKN includes human, social and organizational factors, as well as technical factors. Creating the OKN requires deep engagement among domain knowledge experts and a variety of other stakeholders. These include data owners, decision-makers, various end-user communities, tool builders, and data visualization experts.

Launching an Innovation Sprint

Reacting to the various drivers for the OKN, the NSF, in partnership with OSTP, launched an OKN Innovation Sprint to bring together a diverse set of stakeholders to help design a roadmap for a prototype of an OKN (Proto-OKN). Running from February through June 2022, the Innovation Sprint harnessed the collective insights of experts from over 24 academic institutions, 15 Federal agencies, 3 non-profit organizations, and 20 private-sector companies.

Experience from the NSF Convergence Accelerator had made clear that user-centered design approaches, stakeholder involvement and alignment, and customer engagement were essential to ensuring design of an ethical, impactful and, ultimately, sustainable OKN. Thus, the Sprint activities focused on approaching the vision for an OKN through end users and uses of an OKN. The Innovation Sprint concluded that building the OKN would bring significant fiscal and temporal savings as effort and expense for data identification and data use continue to rise. The OKN infrastructure would bolster data-driven decision making, reduce the likelihood of using inaccurate and/or incomplete data, support coherent management and access to data resources, reduce expense for development of standards including ensuring ethical compliance, and enhance service provision for a broad range of use cases.

This report summarizes the main findings of the Innovation Sprint.

2. Origins of the OKN concept

The vision of the OKN has been developing robustly and with broad support over the past five years. The initial vision was put forward in 2017 as part of the **NSF Harnessing the Data Revolution Big Idea**. The idea was further developed via community meetings, culminating in a workshop organized by the Federal **Interagency Big Data Working Group** (**NITRD OKN workshop**). The imperative for an OKN to support the AI research environment was also noted in the **final report** of the **National Security Commission on Artificial Intelligence**.

In March 2019, the National Science Foundation issued a **Dear Colleague Letter (DCL)** announcing the Open Knowledge Network as one of the inaugural Tracks of the **NSF Convergence Accelerator**. The NSF objective was to engage multidisciplinary, multi-institutional teams to help identify development paths for an OKN, with a particular focus on exploiting publicly available U.S. government and similar public datasets.

In September 2019, the NSF Convergence Accelerator selected 21 multidisciplinary projects for a one-year **Phase 1** effort as part of its inaugural OKN Track (Track A). Phase 1 efforts included developing the team's idea into a proof of concept, identifying new team members and partners, and participating in the innovation curriculum. At the end of Phase 1, teams participated in a formal proposal and pitch, which was used to select teams for Phase 2. In September 2020, 5 of these 21 projects were selected for a two-year Phase 2 effort. Phase 2 efforts focused on high-impact deliverables and sustainability. Teams transformed their prototype into solutions and developed sustainability plans to continue impact beyond NSF support. For more details on the Convergence Accelerator OKN Phase 2 projects, please refer to Appendix B.

The OKN Track A projects leveraged and benefited from prior work in knowledge graphs and related technologies, including **Linked Open Data** and the **Semantic Web**. However, they also uncovered numerous new technical challenges as well as sociotechnical considerations. In November 2021, the NSF issued a Dear Colleague Letter focusing on the research issues uncovered by OKN efforts (**CISE OKN DCL**).

These prior efforts provided the context for the OKN Innovation Sprint. This report represents the collective findings and recommendations of the Sprint contributors. It serves as a guide for future, iterative Proto-OKN development.

OKN Innovation Sprint

The goal of the OKN Innovation Sprint was to assess the opportunity space for creating an OKN and the possible challenges in its formation. The Sprint kickoff meeting on February 23-25, 2022, attracted over 150 attendees representing all sectors — government, academia, industry, and nonprofits. Community members signed up for the 4-month sprint exercise and organized themselves into 17 different groups representing a wide range of application areas and themes. **Figure 1** displays Sprint activities from the kick-off meeting to the Sprint's conclusion.

Each group identified essential characteristics and components of an OKN and then created specific use cases to help guide its further development. Each group consisted of two self-volunteered group leads, several members, and an active Slack channel. Groups met weekly from March through June 2022, and monthly with the Innovation Sprint Organizing Committee. The use cases resulting from their weekly design efforts are detailed in Appendix A of this report.

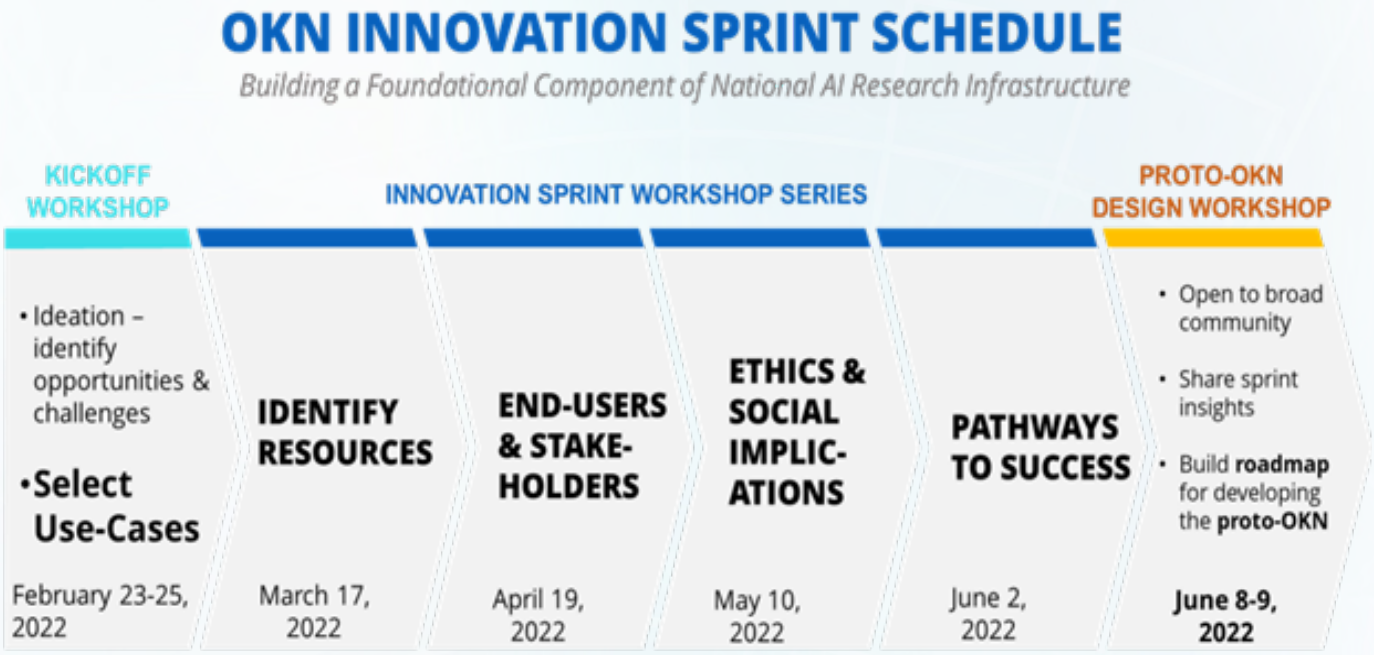


Figure 1: Sprint schedule steps, by date

The Sprint activity concluded with a capstone Proto-OKN Design Workshop on June 8-9, 2022. Each group presented results from their Sprint activities and outputs from all groups were pooled together to serve as the basis for the Proto-OKN design. The OKN Innovation Sprint exercise revealed the tremendous excitement and commitment by the community for the OKN vision and approach and the broad scope of opportunities that it affords.

3. Characteristics of an OKN

The Innovation Sprint revealed the need to engage a broad range of stakeholders and end-users in order to obtain a holistic view of the characteristics of the OKN. Stakeholders ranged from data curators and data programmers to a variety of end users. The Sprint exercise helped capture their needs and requirements as part of the design process.

A user-centered design approach is necessary to engage stakeholders early in the development process. User-centered design shows stakeholders the benefits that they and their clients or customers will receive from the proposed end-product. Designing an OKN that accommodates the needs and preferences of users is essential for its successful adoption. In a project intended to provide easy access to a wide range of users with differing needs, failing to accommodate users' needs will limit the project's success, making it more difficult to sustain over a long period of time.

The Innovation Sprint defined a number of essential characteristics of an OKN, including:

- An open architecture to allow inputs from a variety of sources.
- A dynamic system reflecting real-world information updates and changes as they occur.
- Ability to link disparate information by traversing links across the network and to deduce linkages among entities.
- Interlinking of "horizontal" and "vertical" elements of the network. "Horizontal" elements provide a common technological infrastructure and sociotechnical frameworks, agnostic to the knowledge domains. "Vertical" elements focus on preparation and ingestion of content from specific information domains.

The Innovation Sprint also identified seven elements critical to the establishment and sustainment of an OKN:

- Governance: a set of guiding principles to enable OKN stakeholders to direct and oversee policies, design, development, operations, and management of the system.
- Ethics: structures and processes based on ethical principles that provide consistency, transparency, and accountability.
- Provenance: methods that are able to track any and all changes to information over time, including original sourcing of an asset and subsequent processes that are employed.
- Scalability/Interoperability: a flexible and extensible data infrastructure that is *elastic* and enables the exchange, use, and transfer of OKN resources.
- Sustainability: an extendable ecosystem that enables the OKN to meet current and future needs with backward and forward compatibility.
- Access Rights: a flexible system to enable access to OKN resources from open public access to predefined privileged access and analysis.
- Data Validation: procedures, processes, and metrics for ensuring that ingested data are accurate, complete, and meet specified criteria, including distinguishing trustworthy data from untrustworthy data or misinformation.

An OKN created by integrating a variety of datasets held by diverse entities would:

- Facilitate answering dynamic, interconnected questions for a variety of users, e.g., government employees, companies, private citizens, nonprofits.
- Increase flexibility for ongoing data integration plus future use and data re-use.
- Empower AI and machine learning tools by creating pre-integrated, AI-ready data.
- Enable data-driven insights to tackle the "unknown unknowns" and improve decision-making capabilities for agencies and organizations of all sizes.
- Reduce the expense of data cleaning and "data wrangling."
- Reduce entry hurdles by basing the OKN framework on open, well-established Web standards, allowing use of off-the-shelf, open-source tools.
- Enhance data interoperability by moving from software and platform dependency to semantically interconnected data.

4. Creating an OKN

The OKN is envisioned as an open, domain-agnostic, community-based effort for establishing a national-scale data infrastructure, with the development and maintenance costs being distributed across many stakeholders. The management and maintenance of the OKN is also expected to be a distributed, editorialized process. Established collective intelligence efforts like *Wikipedia* and *Wikidata* provide a model and valuable experience base as a starting point for OKN.

Another important experience base to build upon is the design thinking and user-centered design approaches that have been employed in the NSF Convergence Accelerator Program. These design approaches have proved essential in identifying user needs and requirements, and to develop prototypes and the so-called Minimum Viable Products (MVPs). This approach has not only helped clarify implementation issues and priorities but has also served to engage and empower a diverse set of stakeholders in the process.

The Innovation Sprint identified the activities needed to create a prototype OKN, which include:

- Gathering requirements for use-case capabilities specific to each domain, aggregating capabilities as general requirements by their corresponding horizontal requirements, understanding which of these capabilities are immediately available versus which would require research advances, and identifying the types of linkage required among data from different domains, e.g., health and environment, natural resources, or judicial records.
- Establishing quality schemas in the form of ontologies while taking existing ontologies into account and identifying any gaps, compiling inventories of relevant data resources, services and frameworks across different domains, adopting common or shared representation.
- Encouraging existing repositories to provide easy communication among the prototype-OKN, **NIEM ontologies,** and extensions of the **schema.org** framework wherever possible. Identifying and addressing any barriers to access would help the OKN create more robust data access to various domain repositories, including private and sensitive data.
- Fostering interconnection of information across domains, with particular emphasis on those that may currently be largely disconnected and/or difficult to integrate.
- Developing highly effective prototypes for querying and accessing data and, where applicable, performing reasoning tasks with the data.
- Enabling involvement of domain/subject experts without substantial technical skills, e.g., to verify/validate/curate the knowledge base.
- Prototyping various user-friendly interfaces with different data access modalities that enable expert, non-expert, as well as non-technical users to access and use the data, information, and OKN services.
- Developing metrics to objectively measure the use and the impact of OKN use by different stakeholders.
- Ensuring that it is possible to source data from a wide variety of resources including unstructured, semi-structured, and structured sources and, importantly, data with varying levels of quality and fidelity.
- Developing an approach so that the open system design of the OKN would also be able to support

the ability to incorporate private and access-controlled data using established governance principles and procedures.

The flow of data and information in the OKN system could be considered in several distinct phases: *Design, Ingestion, Enrichment, Storage, Consumption*, and *Maintenance*.

- *Design* would consist of a description of target use cases with requirements and assessment of relevant data sources. An ontology and schema of the knowledge graph for the given use case would then be developed based on these descriptions and assessments. The schema should be developed, documented and made available alongside the graph data to support future updates and evolution of the graph and its use cases.
- *Ingestion* would include setting up data pipelines and workflows to construct/create the graph according to the design. This includes registering and establishing access to external data sources, mapping "legacy data" to the knowledge graph structures, and incorporating data-quality checking measures for each data source.
- *Enrichment* would consist of a set of methods designed to increase quality and applicability of the graph using techniques such as entity disambiguation, aligning/mapping schemas and data to existing ontologies and graphs, incorporating measures to help increase data quality, and making data integration decisions explicit, reusable, and reversible. Enrichment could also include the use of knowledge management methods for completing data-predicting links, discovering links to other OKNs, and assessing data quality. The assessments may include identifying inaccurate and contradictory data, creating derived data, representing different contexts for the data, and obtaining different semantic views to serve different stakeholder/user communities.
- *Storage* would include use of state-of-the-art technologies for managing the ontologies and the knowledge graphs themselves using state-of-the-art storage techniques including graph databases and triple stores using cloud-based implementations. Storage and access via the cloud would help establish OKN as a widely shared infrastructure.
- *Consumption* would typically be via a variety of user interfaces as well as software interfaces, namely application programming interfaces (APIs), for tasks including knowledge graph access, ontology management, provenance/lineage management, and other re-use and administration tasks. User interfaces/APIs would be required for a wide range of functions, including OKN curation, data administration, data exploration and querying of information. Software interfaces (APIs) would be based on open standards (e.g., the W3C "Semantic Web stack" and other related specifications) to allow third parties to build applications using the OKN.
- *Maintenance* of the schema and graph would be secured on an ongoing long-term basis. Suitable versioning (e.g., through distinct releases) would be used to minimize disruptions to applications depending on the graph.

The approach to building the OKN system should be iterative and evolving, to facilitate flexibility and adaptation of the system to changing needs. A set of measurable metrics should be developed to help gauge functionality, performance, and progress. These would be reinforced through continual user testing.

There should be a centralized, lightweight governance board, with a hierarchical structure that enables distributed data governance at the dataset level. The governance structure would include a "data working group" to help interface with federal agencies and other data holders, and to help these stakeholders integrate their data. To ensure an ethical system, the system governance should adhere and adapt to current frameworks and guidelines, including supporting open standards and **FAIR** data (Findable, Accessible, Interoperable, and Reusable) while regularly evaluating the system for potential harms.

5. Considerations moving forward

The OKN represents a best-in-class opportunity to provide **FAIR access** to open data, to enable AI and ML tools and ecosystems, and to leverage data and information needed to address societal challenges and innovation opportunities. The OKN Roadmap provides guidance for how a cross-government effort for Proto-OKN development should proceed. The prototype should strive to provide a public data infrastructure for the use of government agencies and other stakeholders. The prototype would demonstrate the ability to integrate and use public data to develop novel solutions to some of today's most complex challenges facing Americans. It would also serve to enhance public access to data resources created with American taxpayer dollars.

As already mentioned, the OKN would leverage the technological advances and experience base in several areas including the semantic web; ontologies and ontology engineering; the Wikidata ecosystem; Linked Open Data, and others. It would leverage available technologies and existing standards including those from **W3C**, **IEEE**, **OASIS**, and **NIEM**. At the same time, this activity will also help define new standards, processes, and methods in technical areas such as standardizing workflows/processes for data ingestion and for data enrichment, as well as in the sociotechnical aspects of creating a national infrastructure like the OKN.

Strong stakeholder engagement is essential for the success of OKN, and it would be built around the following key considerations:

- *Participatory Design*. OKN would be developed on human-centered design principles that engage stakeholders and end users. Importantly, this would include development of education and training materials that address a broad audience — from technical staff to end users and senior management, to the general public.
- *Open contributions*. To avoid the risk of a closed, proprietary system that may benefit only a few, the system would be open to all. Transparency is an essential attribute of a trustworthy system. The data, metadata, and the processes employed for data ingestion, enrichment and consumption would be open, accessible, and transparent.
- *Ethics*. Ensuring an ethical, responsible system requires an intentional and community-centric approach and careful design, along with ongoing monitoring. Many factors will contribute to the success of this objective, including working directly with communities and ensuring inclusivity in the data, users, and communities engaged.
- *Sustainability*. An open, inclusive system requires a committed community of stakeholders to ensure

sustainability of the purpose, data and network. Starting by interconnecting use-cases of clear value to agencies and other organizations will help build that community, as will connecting to existing open science efforts, data library initiatives, and shared research computing structures. Creation of or inclusion in a coordinating entity such as the proposed **National AI Research Resource** could be valuable for developing the long-term support needed for operation and maintenance of OKN as an infrastructure.

- *Extensibility*. The OKN should be architected using existing standards described above, but with an eye toward the future. New data, use cases, and partnerships must be able to connect to and use the OKN to provide more value and impact.
- *Connectivity*. Maintaining a focus on the sources of data, end-users of OKN tools, and other stakeholders is essential. The human-centered design inherent in developing the use-cases and building the connective fabric of OKN should help prevent fragmentation.

6. Use cases

The OKN Innovation Sprint activities approached the OKN design process through the lens of a variety of use cases, each focusing on a specific societal need. Participants formed 17 use case groups. Thirteen of those groups focused on specific use cases, or "vertical" applications, of the OKN, and four focused on cross-cutting, "horizontal" themes. The 13 "vertical" use cases fall in the broad categories of:

- Equity, Social Care, and Justice Issues
- Climate Change, Disaster, Energy Systems
- Health Communications and Information Accuracy
- Innovation and Research Ecosystems, and
- Macro-level issues, including:
  - Supply Chain Information
  - Data-driven Decision Support, and
  - Financial Risk Analysis.

A more detailed discussion of each use case is provided in Appendix A. Each use case in the following narrative bears the name of the group assembled to create it and by the factors necessitating its creation.

Equity, social care, and justice issues

The following three OKN use cases focus on the concept of providing community care to improve community health and well-being for all community members. Community care would be delivered through social services such as decarceration service planning and family reunification; addressing and preventing homelessness; and increasing transparency in the justice system.

Group B: Integrated Justice Platform

This use case group explored the data infrastructure needed to collect, aggregate, and harmonize data across domains of the justice system in order to improve the way it works, reveal patterns and trends across data systems, and evaluate the influence of bias and other extralegal factors on institutions, communities, and individuals.

Group H: Homelessness

This group explored requirements for creating a knowledge graph that would provide near real-time data tracking of homelessness, available housing, and shelter and social services usage. It would also track related funding and programs to help city and community leaders house unhoused citizens and prevent future homelessness.

Group M: Decarceration Service Planning System

This group focused on development of OKN-based social services that would enable persons released from incarceration to successfully return to their communities and families. These services would help reduce recidivism and improve the well-being of formerly incarcerated individuals and their families.

Climate Change, Disaster, Energy Systems

The four use cases in this group are broadly related to climate change and its impact on food systems, various aspects of local communities and community-level decision making, energy systems, and on natural disasters which, due to climate change and other factors, have routinely become compounded events.

Group E: Food and Climate

This group focused on developing a food and climate use case. They determined that applying existing climate and food-system ontologies to public-agency award data would improve investment outcomes and enhance understanding among investment decision-makers. Ontologies could then provide access to more comprehensive information. Decision-makers accessing this information would be better able to make informed decisions about where food investments would provide the greatest benefit. They would also be able to more easily identify investment gaps.

Group G: Climate Change

This group focused on a multi-faceted decision-support system that uses a coherent set of climate data to help local to global community residents and decision-makers make informed decisions. Such decisions have the potential for significant impact on community health and infrastructure, criminal-justice system equity, and environmental protection. Poorly made decisions could harm individuals, communities, and the environment. Decisions made with the support of the proposed OKN-based decision-support system could help. The OKN would mitigate a broad range of issues and enhance resilience by revealing interrelationship among social, economic, and environmental factors.

Group L: Energy

This group explored development of an OKN-based decision-support tool that would enable the Department of Energy, public utilities, and local communities/stakeholders to:

- Assess risks to the grid from extreme events
- Determine community energy resilience, energy justice, and burden, and
- Identify opportunities for renewables to reduce energy burdens while achieving the objectives of clean energy.

Group P: Compounding Disaster Events

This group explored the use of an OKN to aggregate and connect diverse, multi-scale, distributed data. They looked at how a knowledge graph of this kind could deliver prognostic and real-time information to decision-makers before and during disaster events. Such events tend to co-occur, resulting in complex impact scenarios that require coherent responses from a wide range of stakeholders. In evolving, uncertain contexts, poor decisions could be made without knowledge-graph support. And climate-change impacts could be made worse.

Health Communications and Information Accuracy

The use cases for this topic explored how bona fide public health information could be made available to counter the significant amounts of misinformation on this topic in social media, especially focusing on vulnerable populations in this context.

Group F: Health-Related Information

This group explored the development of a data infrastructure for health-related information — as well as misinformation — with the goal of mitigating health misinformation and improving public health. The OKN developed to meet these goals would promote healthy behaviors, especially among vulnerable populations, such as older adults, low-income families, and minority communities.

Innovation and Research Ecosystems

Use cases for this topic explore how to enable data-driven, insightful decision-making support for stakeholders of various research and innovation programs. Specifically, they explored how to enable the system to support researchers seeking to identify current research gaps and opportunities.

Group C: Defense Innovation Programs

This group investigated construction of an OKN-based repository of all innovation programs currently funded by the U.S. Army and other Department of Defense agencies. The proposed repository would help users identify linkages among various programs, eliminate duplication of effort, initiate new collaborations, and accelerate the transition of technology at scale to end users.

Group R: Gaps in Research

This group investigated how development of an OKN that aggregates micro-level data on research activity could help to identify overlooked research approaches; find gaps in research as targets for future work; and facilitate a more equitable and robust allocation of funding and better target human resource training.

Supply Chain, Decision Support, Financial Risk Analysis

These use cases address macro-scale issues in supply chain management, decision support, and financial risk assessment. They promote the use of data-driven insights, decisions, and predictions for use by public and private agents to tackle the issues listed below.

Group A: Strategic Supply Chain

This group considered development of an OKN for several uses including real-time visibility of supply chain flows (e.g., for water, masks, and medical supply); identification and prediction of bottlenecks and shortages; crisis management and disaster recovery; sustainability strategies; and industry health. Supply-chain resilience is essential for each of the above.

Group J: Decision Support for Government Leaders

This group investigated development of an OKN-driven decision-support system at federal, state, and local levels to provide citizens and decision-makers data-driven insights and predictions related to a broad range of decisions. These pertain to the following:

- Land-use
- Transportation options
- Clean air and water regulation, and
- Emergency and disaster response.

Group Q: Public Industry Risk Analysis and Alert System

This group focused on creating a business and finance OKN curated from publicly available sources. The knowledge graph would provide indicators and indices pertaining to opportunities and risk factors. Publicly available sources could include:

- SEC filings
- Analyst reports, news reports, and Web text

Cross-cutting ("Horizontal") Issues

The use cases for this topic explored common technological infrastructures and sociotechnical frameworks that could be used to create the OKN.

Group D: NIEM Ontology

This group explored the use of the **National Information Exchange Model** (NIEM) to build an ontology that can serve as a foundation for other knowledge-graph design efforts and sample knowledge graphs. The ontology would help achieve faster cross-domain interoperability by leveraging the underlying ontology. This would improve integration and interconnectedness among knowledge graphs/networks across a broad variety of scientific domains and governmental entities.

Group I: Electronic Consent Services

This group investigated the possibility of developing an electronic consent service as a common mechanism for managing privacy preferences. The consent service would cover consent assertions for protecting personal and/or sensitive information in an OKN-based data-sharing system.

Group K: Collaborative Knowledge Graph for Researchers

This group explored a distributed software environment for the +OKN where participants could establish their own "nodes" or leverage pre-existing infrastructure to publish, access, and consume knowledge graphs with privacy and ethical safeguards in place.

Group N: Learning Resources

This group focused on the needs and requirements for an OKN Training and Education Gateway to provide learning and outreach materials to the entire spectrum of stakeholders from agency administrators and elected officials, to technical/software architects, data contributors, professionals, educators, researchers, students, and members of the public.

7. Schedule and Effort

The OKN effort should leverage existing work in related areas and the current technological base. The use of user-centered design methods and iterative implementation methodology will be important in order to match evolving needs and requirements of various use cases. Prior efforts like the NSF Convergence Accelerator OKN Phase 2 and Wikidata should be recognized and leveraged, while also paying attention to the significant new needs and requirements uncovered during the Innovation Sprint. The schedule should take these factors into account while developing the goals, objectives, and deliverables for the short, medium, and longer term (i.e., over 6 months, 2 years, and 5 years). **Figure 2**, below, displays the goals for years 2023 through 2028.

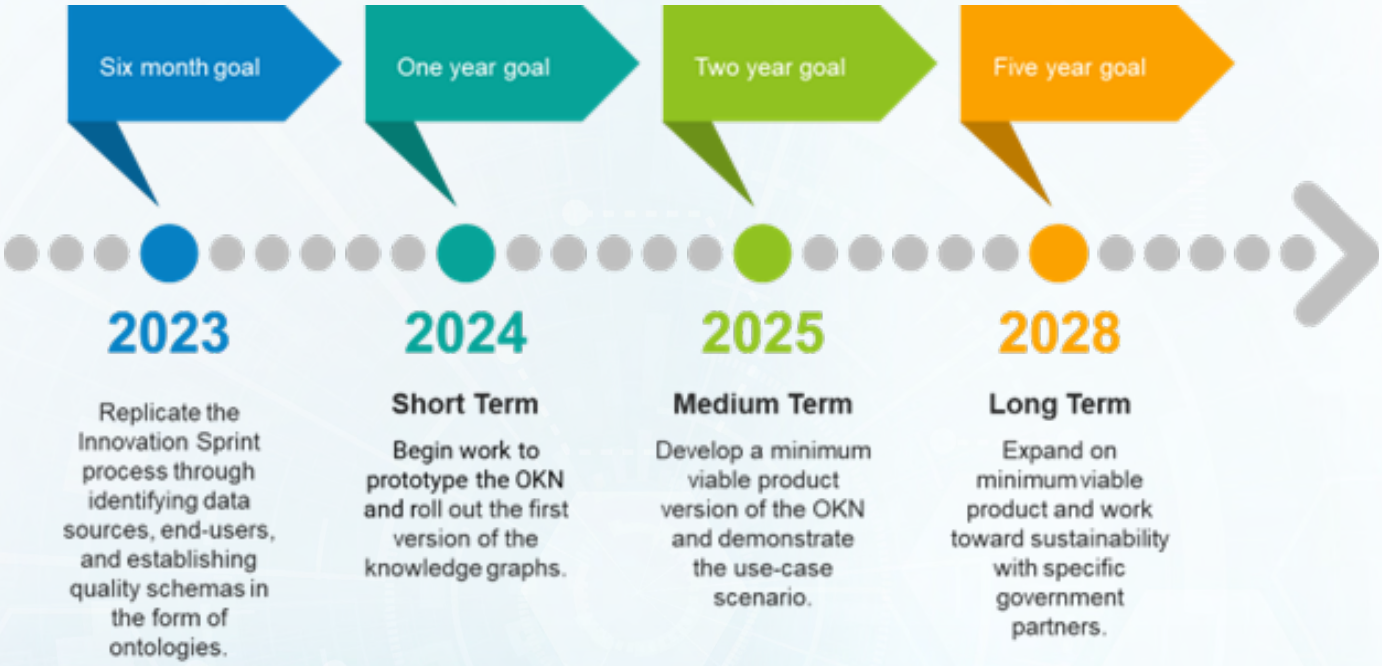


Figure 2: Four phases of OKN development over 5 years

8. Conclusion

Creating the OKN as a network of interconnected knowledge graphs will power the next data revolution by enabling the use of vast troves of data and information to:

- Develop and share real-world knowledge
- Accelerate collaborative innovation, and
- Address current and future societal challenges

The ever-increasing generation of digital data and the need to integrate these data for a broad range of applications with societal impact requires a platform like the OKN. Knowledge representation at the scale envisioned by OKN is essential to the success of future AI-based methods. As a national-scale infrastructure the OKN can benefit a broad range of constituents including government agencies, private organizations, non-profits, academics, and others. It can drive a broad range of novel applications aimed at tackling societal challenges, including use of open data for evidence-based policies, and also help in developing novel AI capabilities.

Publicly supported OKN-related activities have made significant progress over the past few years beginning with Track A in the **NSF Convergence Accelerator** (see Appendix B). The OKN Innovation Sprint has generated considerable community interest and enthusiasm because the OKN holds tremendous potential to impact transformative change not only within individual sectors and domains, but for the nation as a whole. This is the opportune time to embark upon the creation of OKN, beginning with a Proto-OKN development activity.

References

Baxter, G. and Sommerville, I., Socio-technical systems: From design methods to systems engineering, *Interacting with Computers*, Vol. 23, No. 1, January 2011, pp. 4–17.

Berners-Lee, T., Hendler, J., and Lassila, O., The Semantic Web, *Scientific American*, Vol. 284, No. 5, May 2001, pp. 34-43.

Chaudri, V., Baru, C., Chittar, N., Dong, X., Genesereth, M., Hendler, J., Kalyanpur, A., Lenat, D., Sequeda, J., Vrandečić, D., and Wang, K., Knowledge Graphs: Introduction, History and Perspectives. *AI Magazine*, Vol. 43, No. 1, March 2022, pp. 17-29.

Guiterrez, C. and Sequeda, J., Knowledge Graphs. *Communications of the ACM*, **Vol. 64, No. 3, March 2021, pp. 96-104**.

Hogan, A., Guiterrez, C., Cochcz, M., de Melo, G., Kirranc, S., Pollcrcs, A., Navigli, R., Ngonga Ngomo, A-C., Rashid, S. M., Schmclzciscn, L., Staab, S., Blomqvist, E., d/Amato, C., Labra Gayo, J. E., Ncumaicr, S., Rula, A., Scqucda, J. and Zimmermann, A., **Knowledge Graphs**. SpringerLink, 2022.

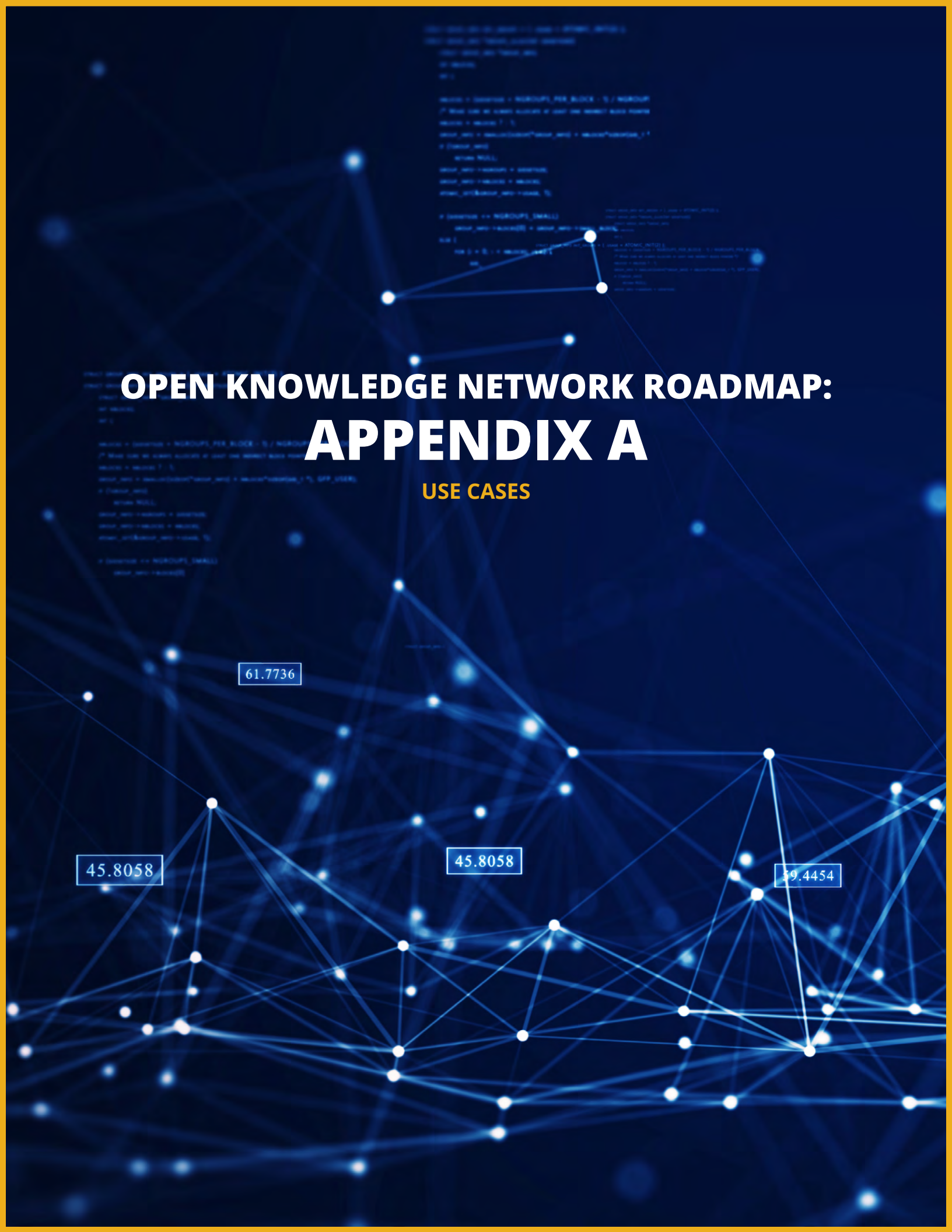
Hitzler, P., A Review of the Semantic Web Field, *Communications of the ACM*, Vol. 64, No. 2, February 2021, pp. 76-82.

National Science Foundation (NSF), "**Harnessing the Data Revolution at NSF**", 2020.

National Security Commission of Artificial Intelligence (NSCAI). **The Final Report**, 2021.

Noy, N., Gao, Y., Jain, A., Narayan, A., Patterson and A. Taylor, J., Industry-scale knowledge graphs: lessons and challenges, *Communications of the ACM*, Vol. 62 No. 8, July 2019, pp. 36-43.


36.7702
30.7317
39.4454



OPEN KNOWLEDGE NETWORK ROADMAP:

APPENDIX A

USE CASES


61.7736
45.8058

TABLE OF CONTENTS

PROTO-OKN USE CASES:

THE FOUNDATION OF THE PROTO-OKN ROADMAP REPORT

Appendix A describes the use cases developed by the participants of the Open Knowledge Network Innovation Sprint, held March through June 2022. These use cases serve as the foundation of the proposed Open Knowledge Network, intended to provide easy access to data now stored in siloed databases.

Each use case targets a specific threat to the environment or to the health and well-being of the U.S. population. Use-case Groups propose to build a knowledge graph to help mitigate specific threats. Knowledge graphs will become components of the proposed Proto-OKN, which will be an interconnected network of individual knowledge graphs. This will provide easy access to users seeking specific information for use in threat mitigation. It will also enable users to share data now residing in siloed databases.

Groups assembled by the problem area that they volunteered to address. Each of the 17 groups tackled a specific problem area, separately identified by group name. All groups used prepared use-case templates to develop their use cases. The results of each group's efforts are provided in this appendix.

Group A: Strategic Supply Chain Resiliency

Use Case A was selected for development based on recent disruptions to the U.S. Supply Chain. Group A contributors are volunteers representing key stakeholders from academia, government, and industry. Stakeholders have direct experience with or interest in this issue.

Contributors

Group A contributors are listed on the following page, along with their organization.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Cavan Paul Capps	U.S. Census Bureau
Vinay Chaudhri	JPMorgan Chase & Co.
Melissa Cragin	University of California, San Diego
Ying Ding	The University of Texas at Austin
Jennifer Dunn	Northwestern University
Jianxi Gao	Rensselaer Polytechnic Institute
Jeff Heflin	Lehigh University
Michael Moore	Neo4J Graph Technology
Emanuel Sallinger	Vienna University of Technology and Oxford University
Dalia Varanka	U.S. Geological Survey
Krzystof Janowicz	University of California, Santa Barbara
Matthew Lange	IC Foods
Ellie Young	Common Action

Champion

Ellie Young, founder of Common Action, was champion for Group A. As champion, Ms. Young organized use-case group meetings and briefed contributors on group updates.

How supply chain resiliency serves the public good

After years of disruption, the inner workings of the American supply chain have become a matter of strategic national priority. Successful delivery of PPE, baby formula, and other essential goods requires successful coordination between individual companies and operators along a vast network of ports, roads, railroads, warehouses, and retail locations.

Providing more and better-quality information across the supply chain and its many hubs across the globe will increase supply chain efficiency. Goods such as medicines and manufacturing components will be delivered as needed, improving health care and reducing shortages.
Hub operators and companies awaiting goods struggle with a continuous mismatch between resource availability and need. At root, many of the challenges faced by operators are informational. Limited visibility of shortages and bottlenecks and a lack of communication between supply chain participants limits their ability to make coherent decisions.

These issues result in recurrent expensive delays, unpaid dwell times at bottlenecks, lack of available transportation, limited storage capacity, and shortages of equipment, parts, and labor. With a deficit of real-time information across the market at large, even tracking resources — from raw materials to manufacturing parts — is a major challenge across the entire supply chain. Without real-time information, decision-makers are limited in their capacity to assess risks and opportunities, and thus to achieve improvements in supply-chain functioning.

Lack of contextual information

For example, although each supply-chain participant is fundamentally influenced by the activities of numerous companies and logistics operators — in addition to contextual conditions such as weather, infrastructure, and fuel conditions — each operator receives little information from peers upstream, downstream, and across the network. As a result, decision–makers must route resources without visibility into real–time conditions. The resulting responses are disjointed. Faulty decisions perpetuate disruption patterns, resulting in cascading failures that reverberate throughout the supply chain.

Lack of coordination

Although supply-chain disruptions are common, operators can only respond to them individually. For example, any given logistics company can only orchestrate their own fleet of trucks. Thus, some percentage of trucks will run with empty loads – even as demand for trucking, on the whole, exceeds supply. However, if it were possible to orchestrate demand at the level of the national fleet, truckers would be free to address the needs of the entire supply chain at large, such that an independent trucker could look for work based on dynamic availability and real-time demand.

Resolving supply-chain disruptions thus requires enhanced access to contextual, real-time information for individual decision-makers, as well as the ability to coordinate across the entire inventory of resources (e.g., chassis, trucks, railroads, and warehouse space) available in the supply-chain system.

Cost of discovery

Even identifying all the factors that contribute to supply-chain problems is a challenge. Although disparate supply-chain participants share some similar data requirements, there is no common mechanism for data discovery. Thus, each company must bear the costs to identify, locate, and procure necessary information and define strategy. Simultaneously, rising costs and constrained budgets place additional restrictions on finding needed solutions to supply-chain challenges, with implications for the entire economy and consumers across the nation.

An Open Knowledge Network

Each of these problems could be solved with a dedicated technology platform that provides a shared, nonproprietary space for access to general-purpose data — such as weather conditions — and to special-purpose information, including reporting and communications between coordinating partners and system-wide reporting and orchestration.

Such a platform could provide a single access point for terabytes of essential data. It could also provide a portal for supply-chain participants. Through the portal, participants could share and transact data and perform services, such as distributing products. This platform would provide an immediate option for all supply-chain participants to identify contextual information relevant to their immediate decisions. It would also help them identify dynamic opportunities for response. Ultimately, it would enhance their problem-solving capability.

At the system level, such a platform could provide a foundation to coordinate responses and actions between operators at local, regional, and national levels. Each link in the supply chain could be equipped to contact and communicate with one another to solve multi-stakeholder problems in concert. Furthermore, events occurring in different locations and sections of the supply chain could be crowdsourced from proximate individuals. That would equip actors to incorporate into their decision-making processes information at broader geographic and temporal scales.

An open knowledge network can provide a foundation for such a system, equipping stakeholders to trade essential, real-time information in order to anticipate, coordinate, communicate, and resolve dynamic logistical challenges.

Goals & objectives for the Strategic Supply Chain Resiliency OKN

Group A contributors envision a supply-chain open knowledge network as a digital twin, representing a map of the resources, processes, actions, activities, and participants who together form the output of the supply chain as a whole. One layer will collect confidential business transaction data that map supply-chain activity via “zero trust” confidential computing techniques. These could be used to create

official statistics of economic activity. Another layer would map public activity and public resources. This material, mapped in layered stratum, could then be aggregated to form an integrated digital fabric, capable of connecting people and processes, and providing a mechanism for documenting and sharing real-time actions, activities, and conditions across the entire supply chain. This would facilitate local self-organization across various levels of supply-chain activities and operations. To meet the specific, transactional objectives of the supply chain and to provide incentives for the usage and maintenance of this platform, the OKN could be constructed to also support a marketplace for connecting:

- Logistical suppliers with demand opportunities, and
- Companies with experts who possess special skill sets or knowledge or who administer particular sectors of the supply chain (e.g., ports and railroads).

As a further example, through an OKN marketplace, any individual trucker may advertise their location and travel status, to find just-in-time load opportunities, thus reducing the frequency and number of no-load trips. As they travel, truckers may also share information about landscape conditions, such as disrepaired infrastructure, or the ramifications of disruptions like the Colonial Pipeline hack that resulted in gasoline rations throughout the Southeast for weeks.

To represent the entire suite of challenges in the supply-chain context, each layer of the supply-chain system would be modeled conceptually in the OKN, including:

- Inventory supply chains, trucking routes, and associated resources
- Map location of trucks, rail cars, warehouses and current status of use — via crowdsourced data reporting and mobile marketplace coordination (e.g., location and availability of trucks connected to available loads)
- Inventory supply-chain participants, immediate partners, indirect dependencies, and coordination points between participants
- Defined contextual risk conditions, such as weather, infrastructure disrepair, and traffic
- Tracked origination and travel of goods and resulting prices
- Identified consumer-demand locations and alternatives for essential items (e.g., baby formula), where possible.

This approach to information representation affords dynamic applications. To communicate with their immediate networks, users of the OKN would then be able to map themselves to various types of resources, to share information about the existing status of those assets with individual colleagues or into the common knowledge base. Resources, such as vehicles and warehouse space, can be discovered by location, allowing discovery of essential resources at the point of need. Multiple data types can be combined to plan routes, such as weather conditions, fuel costs, and driver availability.

Users may also share and avail themselves of public access contextual resources, which can be absorbed into parallel private OKN networks. Real-time information about the status of assets is provided dynamically throughout the system as a component of value-driven workflows for participants. Additional value propositions include:

- Near-real time visibility to supply chain flows for strategic commodities and goods
- Identified and predict bottlenecks and shortages
- Reduced inflation
- Facilitated risk assessment and planning, crisis management, and disaster recovery
- Enhanced sustainability
- Facilitated access to upstream/downstream situational information and analysis of risk for companies
- Enabling user groups to coordinate network-based logistics initiatives (such as load allocation across a national fleet of drivers) and multi-modal route optimization, including analysis of fuel costs and sustainability objectives
- Enabling user groups to collectively evaluate the health of their industry and benchmark efforts against peers-to achieve supply chain resilience.

Beyond the immediate operations of the supply chain, reported data may also be leveraged to provide inputs for valuable analytical processes conducted by a wide range of organizations, including:

- Federal agencies:
  - The U.S. Department of Transportation may determine where to make infrastructure investments, based on reported infrastructure conditions
  - The U.S. Department of Energy may work directly with investors and manufacturers to develop clean- energy supply chains
- Private-sector companies, which can benchmark against peers and also gain access to data to inform and environmental, social, and government (ESG) strategy and other business decisions
- Nonprofit organizations, which can identify a wide range of needs that they can support across a wide range of topics affecting operations of the supply chain, the broader economy, workforce, environment, sustainability, and so on
- State governments, which can identify investment requirements (e.g., in infrastructure and education), and regulatory opportunities

Lastly, this system would provide a fundamental data architecture for responding to future disruptions, such as natural disasters, resource scarcity, and other challenges produced in a world facing climate change, warfare, and unrest.

Resources for Use Case A

Use Case A identifies existing resources that could be incorporated into the OKN. These include relevant datasets, knowledge graphs, ontologies and/or AI/ML/DL.

It also identifies resources that do not exist and would need to be created.

Existing data resources

The list below identifies existing data resources for development of the Strategic Supply Chain Resiliency OKN.

Datasets

This list of datasets below exist and could be obtained for use in OKN development.

- List of national-interest supply items (e.g., N95 masks, bottled water, petroleum, computer chips, and baby food)
- Supply-chain participants
- Private supply-chain transaction data from industry, including shipment data from retail and wholesale, carrier transportation data (e.g., which routes are used and any transportation-mode changes(e.g., moving goods from rail to truck to fulfillment centers)
- Warehousing and storage data (perhaps including inventories)
- Stock sales reports from oil and gas companies and the spot price for oil and gas.
- Resource data to predict specific risk (e.g., U.S. EPA Safe Drinking Water Information System, U.S. EPA Clean Water Needs Survey database; Climate/weather data (dictates storage vs. distribution of energy carriers)
- Crowdsourced real-time resource mapping
- Emerging: blockchain, Bluetooth

Knowledge graphs

Current knowledge graphs that could be used in OKN development are listed below. With private-sector permission, these graphs could be obtained for use in OKN development.

- Private-company knowledge graphs, such as those created and used by financial companies
- Knowledge graphs developed to model particular contextual factors relevant to supply chain operations. These include:
  - **KnowWhereGraph** for disaster-related context and geographical features
  - **Urban Flooding OKN** for flood disruption information)

Ontologies

With the permission of their owners, the ontologies listed below could be used for OKN development.

- Ontology of 22,000 industries within the global economy (hosted by **Industry Knowledge Graph**)
- **Common Action Sustainability Ontology**

AI/ML/DL

AI/ML/DL resources for use in OKN development may be identified at a later date.

Other expertise

The list below identifies additional expertise relevant to the impact and scope of the OKN. Expertise in the areas of social science and user experience research and design would address the needs and preferences of end-users in the context of OKN use and effectiveness.

- Quantitative data collection and statistical tabulation of shipment values, products, weights, mileage, transportation routes, and infrastructure used (rails, bridges, highways)
- Differential Privacy expertise to add noise to tabulated aggregate data in order to ensure that aggregate data cannot be reverse engineered to discover private data Secure computing expertise that would enable "zero trust" collection and statistical tabulation of the data using techniques such as the following:
  - Multi-Party Computing (MPC)
  - Trusted Execution Environment (TEE) computing.
- Qualitative data collection including focus group facilitation and management for identifying stakeholder needs.
- User Experience design and research to:
  - Identify user needs and to create optimal user interfaces
  - Design for marketplace needs
- Incentivize data-sharing incentives
- Expertise related to supply-chain and economic system management. Such expertise is needed to identify common issues in supply-chain management that would benefit from data sharing.
- Expertise regarding how federal, state, and local agencies work with information to manage disasters
- Coordination with agencies to connect supply-chain resource to particular goals

End-users and stakeholders

This report, comprising eight use cases developed by government agencies and private-sector representatives, explores how the OKN will be utilized to serve the public good. It also identifies potential users and stakeholders, listed separately for each of the eight use cases. The list below identifies the users and stakeholders for Group A, Strategic Supply Chain Resilience.

Users

A diverse range of supply-chain participants who would use the OKN to orchestrate the operations activity of the supply chain. These participants include:

- Port, warehouse, truck, and railroad workers
- Third-party logistics operators
  - Shippers
  - Investors
  - Consumers
  - Importer/exporters
  - Retailers
  - Federal agencies affected by supply-chain disruptions): t
- Investors: to anticipate economic shifts and allocate investments accordingly

- Nonprofits: to identify areas for service delivery
- Federal agencies: to conduct business related to management and procurement of supply chain, inflation, and so on
- Demand-side representatives, such as city representatives, public health agencies, consumer representatives, etc., to register consumer needs and coordinate with others to address them

Stakeholders

The list below identifies stakeholders who must be engaged to ensure the development and deployment of the OKN.

Companies

Private-sector companies must be engaged to ensure that the OKN meets their needs and also contributes to its overall effectiveness. Private-sector companies needed to provide essential input to OKN development include:

- Shippers: wholesale and retail companies that ship and order goods
- Carriers: Companies that move goods from one place to another.
  - Major carriers like UPS and FedEx
  - Independent truckers
  - Rail companies
  - Barge companies
  - Airlines
  - Warehousing companies
  - Third-party logistics companies
- Port authorities
- Investors
- Retailers
- Primary consumers/demand-side representatives, such as city representatives, public health agencies, and consumer representatives, etc.

Economic stakeholders

Economic stakeholders are those who will invest in the development and deployment of the OKN. Investments may include financial resources as well as shared data and other information inputs.

Economic stakeholders include the following government policy makers:

- The Federal Reserve
- The U.S. Department of Transportation
- The U.S. Postal Service
- State and municipal planning agencies,
- Federal and local statistical agencies
- Private banks and investment companies, such as JP. Morgan Chase

- Other actors interested in the economic and business activity of the United States

Non-economic stakeholders

Non-economic stakeholders are typically governmental agencies that need to identify resources in times of crisis and map them to communities in need. Crises include floods, tornadoes, wildfires, health emergencies, and pandemics. If natural resources were cataloged in the OKN, long-term plans for drought could be developed. . Plans could also be made and needs assessments conducted for the protection of public lands, and forests.

More importantly, data providers must be key-stakeholders. Specifically, they must have a direct need for supply-chain resilience and a commitment both to contribute to the development of the OKN and to participate in its ongoing operation after its deployment. Data providers must commit to updating OKN data regularly to ensure its accuracy and relevance. To support their commitment and encourage their participation, the OKN will provide as an incentive to key/corporate stakeholders data that is useful to them.

For stakeholders who provide confidential data, it is critical that contributors to Group A understand what industry-level actors might gain from more timely and accurate industry-level statistics. Contributors to this use case assume that for most corporations, the data of individual companies must remain confidential. That is because public access would compromise companies' trade secrets and ability to compete.

For this reason, focus groups must be assembled from the supply-chain ecosystem, which includes industry associations and private-sector businesses. Focus groups would explore the kinds of information that could be gleaned from the industry-wide supply processes to improve individual business operations (see **figure 1**).

NOTE: Currently, many industry associations conduct their own surveys because, to date, federal surveys have not been released in a timely fashion. Consequently, survey results are out-of-date upon release. Results do not reflect current business conditions and thus are of no value to industry associations Also, the questions asked in federal statistical surveys do not reflect the concerns of business. Inviting businesses to participate in data design may prompt increased industry association participation as well as improve the timeliness and geographic and subject granularity of federal surveys.

An outline of the kinds of stakeholders who would be useful in focus group discussions are those who have been active in the Whitehouse **Freight Logistics Optimization Works (FLOW)** initiative. FLOW is an information-sharing initiative to pilot key freight information exchange among supply-chain entities responsible for the movement of goods.

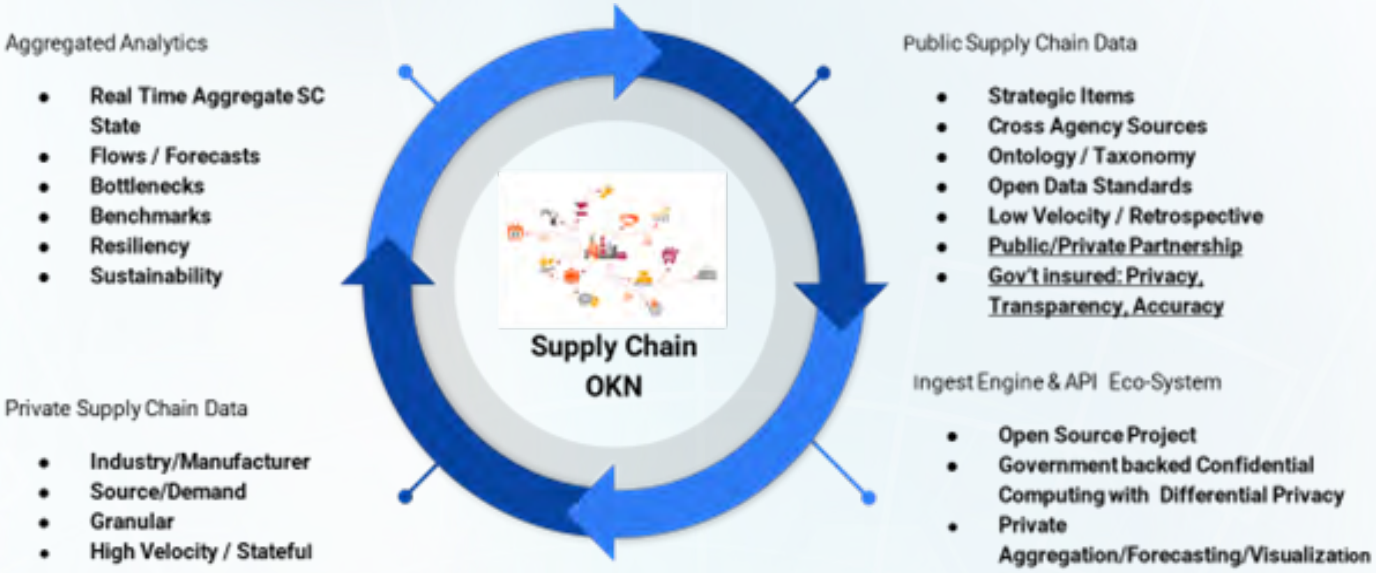


Figure 1: Incentives Flywheel - how the right mix of supply-chain analytics and confidentiality guarantees can incentivize private sector companies to contribute data to the OKN.

Initial participants in the Whitehouse Flow Initiative include:

- **Port Authorities**, including:
  - Port of Long Beach
  - Port of Los Angeles
  - Georgia Ports Authorities
- **Terminal Operators**, including:
  - Fenix Marine Terminal
  - Global ContainerTerminals
- **Chassis**, including:
  - DCLIP
  - FlexiVan
- **Shippers**, including:
  - Albertsons
  - Gemini Shippers
  - Land O' Lakes
  - Target
  - True Value
- **Carriers, Logistics & Warehousing,** including:
  - FedEx
  - Prologis
  - UPS
  - CH Robinson

Ethics and social implications

The **Outcome Driven Innovation (ODI) framework** guided the development of this and all other use cases detailed in this report. Guidance included questions relevant to the implementation and use of an OKN. Questions covered a range of issues, including limitations of data sources, data ownership, as legally defined; core purpose, negative impacts, and potential harms. A **Data Ethics Canvas** was also used to develop this use case.

Solutions specific to the ethics and social impacts of an OKN implementation are listed below.

Semantic ontology

A semantic ontology that describes geography, time, and concepts needed for data discovery would be used to create and implement this knowledge graph. Data would be accessed across various cloud-based nodes via a published API. Only data n published for public access would be available.

Zero-trust system

Confidential data would be invisible to data nodes, to the agencies responsible for creating valid statistical data, and to the owners of the system, using confidential computing techniques like multi-party computing and trusted-enclave computing. There would be no "trusted third-party" computation. This would be a zero-trust system. Data that data providers stipulate as public would be available to the OKN. Statistical summary data processed through formal privacy methods would also be accessible.

Access and transaction controls

Care would also be required in designing controls for accessing data and transacting with other stakeholders in the system/marketplace. The ability to communicate would be controlled by users themselves. In the case of a marketplace where financial exchanges are made, PII would be required for individual accounts, which must be protected.

Relationship to national security

As a representation of the U.S. supply-chain system, unauthorized access to system data could be a threat to national security. System data therefore must be protected from malicious political or commercial actors.

Privacy setting control

As this data is primarily related to business objectives, there is little risk for private trust or abuse. Users will have control over privacy settings and data will be used only for intended purposes that users initiate.

Connecting, sustaining, and scaling OKN

This section identifies the means by which knowledge graphs can be connected, sustained, and scaled within the OKN.

Connectivity

Connecting use cases that affect or are affected by the Strategic Supply Chain Resilience knowledge graph supports cross-connections among OKN nodes, providing users with quick access to additional relevant information. Use Case A connects with other use cases developed in the Proto-OKN Roadmap Sprint. Instantiations of these connected use cases, listed below, provide specific information to the users seeking it.

Use Case Q — Public Risk Analysis and Alert System — Powered by a Financial OKN. This financial OKN is a raw source of data about economic conditions "on-the-ground." Instantiation connects and provides access to business information for a variety of consumers.

Use Case P — Compounding Disasters / Events (a.k.a. Wicked Problems Institute) identifies sources of supply-chain disruption that users in government and industry n must respond to

Use Case G — Enabling Community Climate Action and Local-Federal Intelligence Transmission and Use Case L — OKN for Energy Management enable community and government agencies to access and apply supply-chain capabilities to the green energy climate transition.

Sustainability and Scalability

To be effective over time, the OKN must be sustainable and scalable. This section describes the means by which the OKN will be both.
Sustainability will depend on appealing to the enlightened self-interests of the actors in the supply-chain/logistics space. Allowing companies to see the supply-chain dynamics outside of their own company's control gives them a competitive advantage. Identifying resources useful to current and future markets creates a market for these resources and facilitates keeping resource lists up-to-date.

If enough actors in this space participate, the government could facilitate the use of this network through government contracts or other incentives. Exploring how commercial markets use this information could facilitate public/private support. Focus groups should be tasked with understanding how interdependencies in the logistics market can improve business for all participants.

The budget for sustainability should be approximately $4 million for a two-year effort. Follow-on budgets should be developed based on the results of the two-year pilot. Focus groups must be recruited and assembled to ensure interest and participation by data providers. The value of data collected from focus groups should be analyzed to identify potential new inputs for sustainability. Funding must cover simple multi-party computation processed inside a Trusted Execution Environment (TEE) that includes simple univariate statistics and differential privacy processes used in the publication of anonymized aggregate data.

The ontology that maps geography to time and other topics/concepts should be developed in parallel. An iterative process that verifies the semantics map to real-world practice should be used.

The OKN must be massively scalable because the nodes will reside on the cloud. If the pilot OKN is funded, stakeholders with NSF backing should approach competing cloud providers for support in pilot development. Stakeholders should also approach federal statistical agencies and economic policy agencies like the Federal Reserve, the Treasury Department, the Census Bureau, and the Bureau of Economic Analysis.

Group B: Integrated Justice Platform

Use Case B explores justice-related issues, including the need for an integrated justice platform (IJB) and amelioration of known bias in the justice system. Use Case B identifies possible remedies for the high rate of incarceration in the United States and the high percentage of people of color among prison inmates.

Contributors

Contributors come from academia, government agencies, and the private sector. Each brought their unique experience and perspective of the U.S. criminal justice system.

Contributors to Use Case B and their organizations are listed below.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Kat Albrecht	Georgia State University
Luis Amaral	Northwestern University
Alexia Cooper	U.S. Department of Justice
Christopher A. Cotropia	University of Richmond
Adam Pah	Northwestern University
Erica Smith	U.S. Department of Justice
Paul Wormeli	Wormeli Consulting LLC

Champion

Erica Smith of the U.S. Department of Justice was champion for Use Case B. As champion, Ms. Smith organized use case group meetings and briefed contributors on group updates.

Tracking criminal justice information

The United States leads the world with the highest per capita incarceration rate. Survey and administrative data show evidence that persons of color are disproportionately stopped or arrested by police, compared to their percentage of the population. Very little national data exists on state and local

court systems, resulting in a dearth of information on who is prosecuted and convicted, and what kinds of outcomes they experience. All of these things, taken together, point to our collective responsibility to monitor our justice institutions and intervene, when necessary, to improve how they function and affect our communities.

The fundamental impediment to effective monitoring of justice interventions is that the U.S. criminal justice system is incapable of tracking information from arrest to prosecution to court outcome and beyond. This leaves the U.S. at a deficit when attempting to develop evidence-based interventions to effectively promote public safety and address issues of equity within our justice institutions. The need to better understand how the system functions extends beyond incarceration and directly into our local communities, as a significant proportion of formerly incarcerated persons end up unemployed or homeless, and they experience other continuing consequences of their justice system involvement that inevitably contribute to the cyclical nature of a stagnant system.

The extraordinarily fragmented nature of the criminal justice system contributes to the inability to track and monitor outcomes. This fragmentation keeps the public from knowing how decisions made by one justice agency impact the decisions of other institutions of justice. The resulting dearth of useful outcome measures enables justice institutions to remain unaccountable to the public they are charged to serve.

How the justice system responds to reports of criminal activity, from arrest to prosecution to conviction to sentencing, and what influences each of those responses, should be required information at every level of government. The Integrated Justice Platform seeks to establish the data infrastructure to make the necessary connections across data systems within the criminal justice system, to improve the collective knowledge of criminal justice functioning and outcomes and to evaluate the influence of bias and various other extralegal factors such as the social determinants of crime. Correlating the total response of the criminal justice system to crime is essential to reveal inequities and inefficiencies in the administration of justice.

Use case goals & objectives

Group B set the following goals and objectives for Use Case B:

Goal: Collect, aggregate, and harmonize data across the subsystems comprising the justice system, using a standards-based OKN model promulgated at the national, state, and local levels.

Objective 1: Utilize the Integrated Justice Platform OKN to develop data-driven strategies that improve the way the system works and learn how different individuals move through the system.

Objective 2: Analyze data from the Integrated Justice Platform to specifically examine bias and inequality in justice system case processing and other justice outcomes, to improve transparency in the function of these institutions and inform policy interventions.

Objective 3: Build the Integrated Justice Platform OKN with ethics and equity as a central focus, to improve the data products available for research and policy making and to make more transparent the functioning of the U.S. criminal justice system.

Use Case B resources

A large number of criminal justice data sources exist, but they are fractured. This OKN would combine data from emergency dispatch systems, police contact and arrest, courts, and carceral institutions.

1. Computer-Aided Dispatch (CAD) data
    a. structured and unstructured data collected by emergency dispatch and routed to appropriate authorities, including law enforcement, fire and rescue, and emergency medical services
    b. CAD data generally contain records of officer-initiated activity as well as citizen-initiated activity
    c. CAD data will contain records of crime incidents and other calls for service or requests for emergency assistance that are not criminal in nature
    d. Recommendations for CAD-to-CAD interoperability have been drafted by subject matter experts, but to date, no large scale data collection efforts have been undertaken to gather data from multiple sites and examine the comparability of the data over time and across locations.
2. Record Management System (RMS) data
    a. Local law enforcement crime incident data residing in RMS at the agency-level
    b. RMS data serve as the basis for the FBI's National Incident-Based Reporting System (NIBRS) collection, so much is known about the systems used by local agencies to record incident data
    c. The FBI's NIBRS collection can serve as a source of information on crime incidents captured by local RMS.
3. Arrest and booking data
    a. Local law enforcement capture arrest and booking system data; these data are held in RMS or jail management systems at the local level
    b. Law enforcement agencies that participate in NIBRS will have arrest data connected to their crime incident data, as required by the NIBRS technical specification, but booking and related information is not captured in NIBRS.
4. Jail management system data
    a. At the county- and municipal-levels, these systems will contain information on arrestees detained after booking and additional information about short-term pretrial detention and release type
    b. The IJP OKN would need to investigate whether a source of aggregated JMS data exists that could support the integration effort.
5. Pretrial detention data
    a. These data will include pretrial assessment information, including one or more scores for validated risk assessment scales, if release was authorized, the type of release, and any failure data
    b. Pretrial data are collected at the local level
    c. These data have been accessed by data collectors in the past, which supports the idea that the data are available and could be used for the IJP OKN effort.
6. Prosecution case management data

    a. Data on cases reviewed and charges filed by the county prosecutor, including charge enhancements, limited data on prior offenses, restraining orders, drug involvement, and other case-related indicators
    b. To our knowledge, these data are not easily available and are not collected at the state or national levels.
7. Court case management data
    a. Data on charges filed in state court, including those charged in magistrate's court and courts of limited and general jurisdiction
    b. These data should also contain information on court outcomes such as pretrial diversion, conviction, dismissal of charges, etc.
    c. This data source will also supply information on sentencing, for cases that result in conviction
    d. These data have been aggregated by data collectors in the past, at least for large counties in the U.S., which supports the idea that the data are available and could be used for the IJP OKN effort.
8. Data on jail and prison admissions and releases
    a. Data on persons admitted to jail or prison
    b. Data on persons released from jail or prison
    c. The U.S. Department of Justice, through the Bureau of Justice Statistics, collects this type of data through the National Corrections Reporting Program; however, further investigation would be required to determine if those data are appropriate for use by the IJP OKN.
9. Data on justice involvement across jurisdictions
    a. State and Federal criminal history data provide information on arrest, adjudication, and incarceration across the U.S.
    b. Criminal history records provide data on the number of arrests, court outcomes, and incarceration sentences experienced by each individual person, with the ability to examine data within each state and across states
    c. Criminal history records can also provide data on involvement with the Federal justice system
    d. The FBI facilitates access to criminal history data across states through the Interstate Identification Index but is typically restricted to agencies with a criminal justice need to access the data
    e. Each state maintains their own criminal history data repository, to which criminal justice agencies in the state contribute information; these state agencies may be able to facilitate access to state-maintained information for purposes of the IJP OKN.

This OKN can easily be connected to other OKN projects using geographic indicators. The IJP could be usefully connected to other community-oriented OKNs and other sources of social data, such as those collected by the U.S. Census Bureau and the Bureau of Labor Statistics.

This OKN will benefit from other potential OKN projects, such as establishing a NIEM ontology for data exchange and the Decarceration-Reunification project to understand the various trajectories of persons released from prison and how they fare during their return to the community. The potential connection between the Decarceration OKN and the IJP was also noted in several comments from reviewers and collaborators.

Other expertise

This project would be best served by an interdisciplinary team of social scientists, data scientists, legal scholars and community partners. These community partners are essential and should represent end users and stakeholders (from funding institutions who can help promote use of the data product, criminal justice partners who can provide data, and different user groups).

These groups will work together to quantify resource unknowns in a rapidly evolving terrain where more data is becoming available over time.

Suggestions regarding several specific subject matter experts across the various data domains include:

1. Maria Cardiellos — Executive Director, Integrated Justice Information Systems (IJIS) Institute
2. Brian Acken — Public Safety Team Director, SAS Institute
3. Kay Chopard — Executive Director, Kantara Initiative

The development of the IJP OKN would benefit from the input of a variety of stakeholders whose perspectives may not be reflected in the data underlying the network, including: crime victims; victim service providers; social welfare organizations; faith-based organizations that provide community support services (e.g. soup kitchens, homeless shelters), and prison re-entry service providers. This engagement is critical to ensuring stakeholders provide input into the system structure, function, and privacy considerations.

The technical requirements for this to succeed cannot be understated. The project will require, among other positions, database engineers, data scientists with extensive experience in AI and ML techniques and working with unstructured data using tools like **AWS Glue** or **Trifacta**, and criminal justice practitioners and researchers with expertise in the data needed for this effort.

Tailoring the IJP OKN to the public good

The IJP OKN will centrally focus on maintaining its integrity and serving the public. To accomplish this, it will have processes for ensuring data quality via national standards for data accuracy. This includes auditing processes, quantifying data uncertainty, and transparency about data limitations.

A key tenant of this philosophy is the consistent creation of clear and accessible documentation about bias at different levels of the data pipeline.

Importantly, the IJP OKN helps to ensure that the platform is tailored toward the public, not just system practitioners that already have some level of data access and insider knowledge.

The IJP OKN provides the opportunity to examine how the criminal justice system performs when confronted with the basic task of processing a criminal case across components of the system. In addition to this fundamental question of what happens with a criminal justice event, the IJP OKN can provide insight into basic issues like the impacts of bias on event processing, and how event

characteristics like weapon use, victim demographics, prior criminal history, etc. affect case outcomes. As systems become more sophisticated and capture even more robust and meaningful indicators, the IJP OKN can grow to allow analyses of the data to become more precise and targeted.

End-users and stakeholders

A variety of users and stakeholders can make use of the IJP. We conceptualize these users/stakeholders at three levels: data providers/creators, data connectors, and data users. Data providers and creators include correctional institutions, governmental entities, courts, law enforcement agencies, and prosecutor offices. Data connectors include survey and database curators and maintainers, including existing criminal justice system service providers. Data users include a wide group of justice practitioners, researchers, and the general public, as well as data providers/creators who can become users of the integrated data across systems.

Human-centered design

To understand what users and stakeholders need, Use Case B contributors conducted a series of interviews. In completing these interviews they found that the delineation between system insiders and system outsiders is murky, such that some criminal justice practitioners have less data access than others. They also found that data that is legally public is not functionally usable even for justice practitioners because of fee-related barriers to re-acquiring the same information over and over again. A big part of the problem is that much of the current system of data transmission seems to rely on pro-bono work and paper records. Use Case B contributors plan to take these key concerns into account when designing the IJP OKN.

Potential users expressed a desire for data that was clear about its limitations and applications. Users also stressed a need for a data tool that would be usable for non-data experts.

Ethics & social implications of an IJP OKN

Criminal justice data is fraught with ethical concerns. Part of the problem with criminal justice data ethics is that most IJP data is already legally public, but not accessible to the public. That does not mean that it is inaccessible at all, but rather that data is currently much more accessible to well-resourced companies and individuals. This means that in the current IJP system, the data of individuals is not necessarily protected. But due to inequities in access to the information, auditing fairness and bias in the overall criminal justice system remains difficult.

The IJP aspires to be more ethical and more transparent at the same time. The IJP OKN will provide a series of anonymized outputs to the general public, increasing the transparency of a currently opaque system. To access more detailed microdata, IJP will require a data plan submission process that allows the IJP OKN team to see how data is being used and for what purpose.

The IJP considers downstream use of the data to be part of the ethical story. That is, IJP data should be used responsibly, with clear articulation of its limitations. In order to do this, the IJP will provide robust

and understandable guidance about bias within data and data limitations.

Path to IJP OKN success

This section identifies the means by which the IJP OKN can achieve success. By connecting with other OKNs pertaining to the needs of offenders and victims alike, the IJP OKN will enable better outcomes for both groups. It will also help law enforcement personnel, social workers, and policy-makers understand the social and economic impacts of criminal justice.

IJP OKN connectivity

The IJP OKN can be built in compliance with other cross-cutting efforts like the NIEM Ontology Project. This means the IJP OKN can be designed with the capability to manage consent, particularly for victim-included data using insights from the Consensus Management OKN.

The IJP OKN can also develop connections with other social services projects, like those described in **Use Case H, Homeless Knowledge Network for Cities** and in **Use Case M, the Decarceration and Unification and Reunification from Prison Incarceration to Community Reintegration**, such that this platform functions as connective tissue between other OKNs, which will let us understand the larger social and economic impacts of criminal justice both inside and outside the carceral system.

IJP OKN sustainability

The first step is to select candidate jurisdictions with community partners and OKN experts to demonstrate the viability of the product. Once these data collection systems are in place for a selection of different jurisdictions, they can be scaled up to cover many other jurisdictions.

In the short term, the IJS platform can be built with self-sustaining pathways. Long-term, the OKN will have to be monitored to include new data sources and adapt to changes in existing underlying data structures.

Other concepts and issues

Partnerships with local jurisdictions and community partners are critical to the success of the IJP OKN, namely that the data is collected and delivered to support the needs of users to analyze and understand the current state of and issues in the criminal justice system. Data must be transformed into information that can guide policy-making efforts and assess whether the criminal justice system is effective and distributing outcomes equitably and fairly.

Much of data collection in criminal justice is approached from the top-down perspective, which places the burden of data contribution on the responding organization or person. The IJP OKN provides the opportunity for a 180-degree pivot to a more grassroots approach, whereby the burden is on the data collection agent to understand and work with the data, agnostic to the source or the structure of the information. We believe this shift in the burden paradigm will have a positive impact on the ability to collect these data and realize the success of the IJP OKN.

Group C: Defense Innovation Programs

Group C developed Use Case C to capture current issues related to defense programs. Contributors from government agencies, academia, and the private sector identified cross cuts or points of integration that a Defense Innovation Programs Knowledge Graph could provide to contributing groups as well as to the public at large.

Contributors

Contributors from academia, government agencies, and business were chosen for their interest in and knowledge of defense innovation programs. They are listed below.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Carl Busart	DEVCOM Army Research Laboratory
Christopher Zember	DEVCOM Army Research Laboratory
Emily Rothenberg	Columbia University
Florence Hudson	Columbia University
He Bai	Oklahoma State University
Jed Pitera	IBM Research
Jemin George	DEVCOM Army Research Laboratory
Jianxi Gao	Rensselaer Polytechnic Institute
Lauren Close	Columbia University
Louiqa Raschid	University of Maryland
Michael Cafarella	Massachusetts Institute of Technology
Murat Omay	U.S Department of Transportation

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Nathan Burkholder	Kitewire - Mobility Software Innovation
Paul Nemirovsky	DMetrics - AI Platform for Business
Ying Ding	The University of Texas at Austin
Zeke Topolosky	DEVCOM Army Research Laboratory

Champion

Jemin George of the U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory was champion for Group C. As champion, Dr. George organized use-case group meetings and briefed contributors on group updates.

Benefits of Defense Innovation Programs to the public good

The commercial market is a key source of innovative solutions to tackle the U.S. military's toughest problems. The U.S. Department of Defense (DoD), especially its service labs, have a vast collection of opportunities to leverage innovation occurring among entrepreneurs and non-traditional vendors. However, currently there exists no coordination of efforts between service labs or agencies to ensure that the efforts and outcome of various innovation programs are well utilized/leveraged across the agencies and there is no duplication of efforts.

We propose to construct a data repository, in the form of an Open Knowledge Network (OKN) of various innovation programs currently funded by the U.S. Army and other DoD agencies. An OKN of this kind would allow the Army to better link various programs and initiate new collaborative efforts among participants. These linkages would yield outcomes greater than the sum of their parts and eliminate duplication of effort. Once successful, a Defense Innovation Program OKN would be expanded to include various programs from other services.

One of the main challenges in constructing such an OKN is the lack of automated processes currently available for scanning through the vast amount of publicly available documents and, without human assistance, extracting from them all vital information. Development of such capabilities would require a common ontology and semantic understanding. Also, since new programs are routinely initiated, there would need to be a mechanism for periodically updating the Defense Innovation knowledge network/ OKN.

Goals & objectives

The objective is to construct an OKN of various innovation programs funded by the U.S. Army. Use Case C contributors hope to develop an automated process to scan through the millions of publicly available documents to extract information essential to the construction of this OKN and to its utility after construction. They would require the automated process to extract information without human assistance. Use Case C contributors also hope to develop a mechanism for periodically updating the OKN with new information as it becomes available.

Resources

This section identifies available and needed resources for the Defense Innovation Program Knowledge Graph.

Existing Datasets

Publicly available dataset obtained from the programs listed below could be used effectively to develop this knowledge graph.

- Small Business Innovation Research (SBIR)
- Small Business Technology Transfer (STTR)
- **xTechSearch**
- **Defense Technical Information Center (DTIC)**
- **Defense Information Systems Agency (DISA)**, and
- Various programs listed in the **MITRE Acquisition in the Digital Age (AiDA)**
- **DoD Innovation Ecosystem**

AI/ML/DL Techniques

The list below identifies AI/ML/DL techniques that would be used to develop this knowledge graph.

- Natural Language Processing and other automated information extraction tools.
- Data Discovery Toolkits like **Amundsen from Lyft**, **Metatron Discovery**, **Superset** or **OHDSI WhiteRabbit**.
- Semantic modeling tools like **WebProtege** or **Fairspace**
- Repositories to store and maintain the knowledge base
- Data mapping and visualization tools

Knowledge discovery

An OKN of various innovation programs would allow the Army S&T community to learn about any particular effort as well as identify related programs. Army Program Managers and Program Officers will use the OKN to better link various programs to other efforts, initiate new collaborative efforts among the participants to yield outcomes greater than the sum of their parts, and to eliminate duplication of efforts.

Resources for developing a minimum viable product

Contributors identified the following resources essential to creating a Minimal Viable Product:

- Data extraction and entity resolution and matching and linkage tools. [scispaCy, Neo4J]
- Use existing ontologies and taxonomies.
- Microsoft Academic Graph or Semantic Scholar or ACM DL may be relevant for entity linkage.

Defense Innovation Programs OKN end-users and stakeholders

Contributors identified probable end-users and stakeholders for this OKN.

Users and stakeholders

This OKN would be useful for various program managers and program officers in identifying various efforts related to their current programs. In addition, this OKN would allow the stakeholders to better link various programs, initiate new collaborative efforts among the participants to yield outcomes greater than the sum of their parts, and to eliminate duplication of efforts.

Stakeholders whose involvement is essential to the development and deployment of the OKN include the following:

- **Academia:**
  - Research Offices and TecTransfer Offices will utilize the OKN to :
    - Identify current research opportunities, appropriate gov't agencies and the right partnership opportunities,
    - Publish their metrics of success for past projects,
    - Manage their research portfolios and align their investment/partnership strategies as they grow.
- **Industry:**
  - For industry R&D organizations and small businesses, the OKN will
    - Offer a side-door to engage DoD for small institutions that cannot compete with big defense contractors
    - Allow a wide variety of participants to publish their metrics of success for past projects,
    - Manage their research portfolios and align their investment/partnership strategies as they grow.

Human-centered design

Human-centered design (HCD) will be employed in OKN development. This will ensure that the OKN user interface (UI) incorporates end-users' requirements, including their wants and needs. Specifically, the UI should provide easy access to the information that users are seeking. OKN UIs should also point to or suggest information that users may not have anticipated, but that will be useful in solving their problems.

With HCD requirements for the Defense Innovation Programs OKN in mind, Use Case C contributors

identified end-users and their needs and wants. End-users include the following:

- **National Security Innovation Network (NSIN)**
  - NSIN Mission: build networks of innovators that generate new solutions to national security problems.
    - OKN will support NSIN's talent management efforts through the National Service Portfolio.
    - OKN will allow NSIN to bring together innovators from defense, academia, and the venture community to solve national security problems.
    - OKN will assist NSIN identify startups emerging from the academic and the venture communities that can address DoD problems in innovative ways.

- **U.S. Army Combat Capabilities Development Command - Army Research Laboratory (ARL) — Strategic Partnership Office**
  - OKN will support ARL Director's Action Group (DAG) to identify various ARL engagements with external partners.
  - OKN will provide visibility into ARL partnerships with **DEVCOM Centers** and **Program Executive Offices**.
  - OKN will assist ARL Plans & Programs Office to target their communication.

- **Defense Technical Information Center (DTIC)**
  - DTIC Mission: Aggregate and fuse science and technology data to rapidly, accurately, and reliably deliver the knowledge needed to develop the next generation of technologies to support our warfighters and help ensure national security.

The OKN will enable DTIC to better support the U.S.D(R&E)'s efforts to mitigate new and emerging threat capabilities, enable affordable or extended capabilities in existing military systems, and develop technology surprise through engineering by:

- Preserving and disseminating the research that led to the technologies our warfighters use today.
- Delivering the tools and collections that empower the research and engineering enterprise to accelerate the development of technologies that will help maintain our nation's technical superiority.
- Stimulating innovation by providing access to DoD-funded research and digital data to the public and industry; and
- Maximizing the value of each dollar the DoD spends through the analysis of funding, work-in-progress, and Independent Research and Development (IR&D) data to identify gaps, challenges and the way forward.

Ethics & social implications

This section identifies ethical principles that would guide development and deployment of the Defense Innovation Programs OKN. It also identifies social implications of this OKN's deployment.

Ethical principles

Contributors identified ethical principles that could guide the development and deployment of this OKN. They include:

- The overall data-ecosystem needs to be confined to current rules, regulations and FAIR (findable, accessible, interoperable, and reusable) practices.
- Make sure the proposed use case adheres to the existing laws and rules
  - If needed, strengthen the existing rules and laws
  - Create oversight and accountability structure
- Ethics analysis: consider whether the discretionary data access sharing/use should proceed
  - OKN community need to develop the ethics framework (consider 9 **VA Ethics Principles for access to and use of Veteran Data**)
  - Create oversight and accountability structure
- Consider layered data model with various access rights

Path to success

This section identifies the means by which the Defense Innovation Programs OKN can achieve success. By connecting with other knowledge graphs pertaining to innovations with industry and government, community organizers, nonprofits, industry, and academia would be better able to collaborate with the Department of Defense in the provision of programs and services.

Connectivity

Connecting use cases that affect or are affected by the Defense Innovation Programs knowledge graph supports cross-connections among OKN nodes, providing users with quick access to relevant information.

As one instantiation of the overall OKN, this use case will serve the public, especially industry and academia, by providing them with a broader visibility into various government innovation programs and government decision-making processes. Greater visibility will allow industry and academia to identify numerous societal challenges currently facing the country.

This OKN would offer a front-door for community organizers, non-profits, industry, and academia to engage the federal government. These users could then easily identify novel research areas, collaborations, and funding opportunities. Connecting this use case with other use case data sets allows government agencies to better identify linkages between various programs and initiate new collaborative efforts among participants. In use, these linkages would yield outcomes greater than the sum of their parts, accelerate the transitioning technology at scale, eliminate duplication of efforts, and save taxpayer dollars.

Group D: NIEM Ontology

Use Case D identifies the uses and benefits that a National Information Exchange Model (NIEM) Ontology would provide to OKN users. Users include federal, state, and local governments; tribal and territorial governments; governments outside the United States; and other public and private organizations in the United States and abroad.

Contributors

Group D contributors have a compelling interest in creating a NIEM ontology as well as deep knowledge to contribute to its development. Use Case D contributors include representatives from academia, government, and the private sector.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Nariman Ammar	The University of Tennessee Health Science Center
Tom Carlson	Georgia Tech Research Institute
Kevin Doubleday	Fluree PBC
Katherine Escobar	U.S. Department of Defense
Nat Fuller	U.S. Department of Defense
Michael Grove	Stardog
Brian Handspicker	Open City Labs
Dave Hardy	U.S. Department of Defense
Jeff Heflin	Lehigh University
Pascal Hitzler	Kansas State University
Sharat Israni	University of California San Francisco
Eric Jahn	Alexandria Consulting

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Ora Lassila	Amazon.com, Inc.
Timothy Lebo	United States Air Force
Chengkai Li	University of Texas at Arlington
Christina Medlin	Georgia Tech Research Institute
Art Murray	Applied Knowledge Sciences
Ron Rudnicki	CUBRC
Lauren Sanders	National Aeronautics and Space Administration
Mark Schildhauer	University of California, Santa Barbara
Greg Seaton	Fluree PBC
Cogan Shimizu	Kansas State University
Karthik Soman	University of California San Francisco
Paul Wormeli	Wormeli Consulting LLC

Champion

Eric Jahn and Tom Carlson were champions for Use Case D. Eric Jahn was project management champion. Tom Carlson was technical champion. Both champions organized use case group meetings and briefed contributors on group updates.

NIEM Ontology serves public good with common vocabulary

Group D proposes building an ontology based on NIEM and sample knowledge graphs/OKN use cases. These sample use cases will be developed for OKN users in:

- Federal, state, and local governments
- Tribal and territorial governments
- Governments outside the United States, and
- Public and private organizations

The NIEM Ontology will enable these OKN users to achieve faster cross-domain interoperability in AI, ML, and decision-making. NIEM has an existing broad-based community and infrastructure that can be easily leveraged to develop the NIEM Ontology. It also has tool sets and applications that support data exchanges that could be used to develop the ontology and support its evolution over time. The NIEM ontology would be comprehensive, adaptable, and easily accessed by OKN users. It would enable efficient information exchange among the OKN community with a semantic library of approximately 13,000 vetted terms and definitions across 17 domains.

A NIEM Ontology will serve the public good by moving from a common vocabulary to a common understanding by:

- Developing base ontology modules and design patterns as ready-to-use foundation for others to develop ontologies as schemas for their knowledge graphs/OKNs,
- Mapping to different ontologies and linking those ontologies to the underlying data
- Defining best practices (including the knowledge to protect sensitive data)
- Developing, maintaining, and evolving knowledge graphs
- Leveraging the underlying methodology to improve integration and interconnectedness between knowledge graphs/networks across scientific domains and governmental entities.

OKNs require ontologies as schemas. Developing ontologies from scratch is very expensive and can result in schemas that are not easily compatible with different OKNs. In practice, using a single ontology as a schema for different and evolving OKNs does not yield desired outcomes. Because the schemas of different OKNs do not match, information exchange is impeded. Creating schemas OKNs with very different use-case requirements is not feasible.

NIEM standards could be used effectively to address these issues. NIEM standards are well-established in the OKN user community and provide a common vocabulary for various OKN user groups. Taking the mature body of NIEM standards and casting them into ontology modules and design patterns would support compatibility and information exchange between and among various OKN knowledge graphs.

If implemented, this approach would:

- Significantly lower the cost of quality schema development
- Ensure a base compatibility on the schema or semantic level, and thus

- Support interoperability between different OKNs

The NIEM standard that is currently expressed in XML is suitable for casting NIEM standards into standard ontology formats.

NIEM Ontology Goals and Objectives

The primary goal of Use Case D is to develop a core NIEM Ontology, consisting of separate but compatible modules, along with use cases and knowledge graphs, and including educational and marketing materials.

Resources

Existing resources for Use Case D include the following:

- NIEM as a data model
- Ontology and knowledge-graph expertise in the working group
- Modular ontology development processes tailored towards obtaining quality, versatile, and highly reusable modules
- Software tool (e.g., for visualization and validation, that can be used in the effort).

Resources that need to be discovered or created include:

- Specific use cases to apply to a NIEM Ontology
- Use cases created in collaboration with other Innovation Sprint efforts
- Pre-existing use cases within NIEM domains

Other expertise

Use Case D identified the following three categories of expertise needed to develop the NIEM Ontology:

1. Expertise with NIEM
   a. Already present in the working group
2. Expertise with ontologies and modeling
   a. Already present in the working group
3. Subject matter experts in different knowledge domains that can serve as application pilots
   a. Partially already present in the working group
   b. Available via other user cases, as the ontology expands to support them

Required resources for a minimum viable product (MVP)

Group D contributors identified resources that would be required to produce an MVP. These include:

- Finalized formal model of prioritized NIEM core terms (in OWL format) with detailed documentation
- Detailed examples use case (i.e., demonstration graph based on the NIEM ontology)
- Resource portal with educational and marketing materials

NIEM Ontology end-users and stakeholders

NIEM has an existing broad-based community and infrastructure, along with tool sets and applications that support data exchanges. Evolving to a NIEM Ontology will provide the OKN community with an established semantic library of approximately 13,000 vetted terms and definitions across 17 domains.

The public good is served by moving beyond messaging to higher-level activities like knowledge-sharing, knowledge-based collaborative decision making, AML/ML incorporation, convergence between OKNs.

Group D contributors identified probable NIEM stakeholders and end-users. They asked detailed questions of each user group and obtained answers useful in NIEM Ontology development.

Stakeholders include knowledge graph vendors, researchers, and practitioners. End-users range from consumers to internal enterprises. These user groups and their responses are listed below.

Stakeholder #1: AWS, Google, Azure (Knowledge Graph software vendors)

Please tell us about your research in this arena - what do you work on, any particulars?

I represent a technology vendor (Amazon Neptune cloud-based graph database service), and our work consists of developing the database service as well as helping our customers build knowledge graph-based solutions.

Which organizations do you already work with in this area?

Our customers come from both private and public sectors and range from very large organizations to startup companies.

Is it all researchers or practitioners as well?

Practitioners, particularly people building KG applications and systems (SMEs and software developers alike).

Who are the end users of your OKN?

Whoever our customers target their systems for. This ranges from consumers to internal enterprise users. In the latter case, we see a lot of systems being built that offer "decision support" (to use the term very loosely).

Do you share information, metadata or data? If so, what kind? How do you share it?

Again, this varies a lot. Use of external (shared) data is popular with folks building enterprise KGs.

Which information, metadata or data do you wish you could collaborate on?

See previous answer. Our customers are often interested in building their own ontologies and we typically recommend they rather extend existing ones. From that standpoint, NIEM has the possibility of being a useful "upper ontology".

Which organizations do you wish you could collaborate with more, and for what purpose?

Purely hypothetically, organizations that publish data, to ensure that public data sets are more widely available.

We are thinking about developing a multidisciplinary portal and community to bring together OKN researchers and users, data connections, metadata and communities. What would be most helpful for you from this perspective?

Material (e.g., examples) about how to use NIEM and, more generally, how to use upper ontologies to build your own (domain-specific) ontology.

What else would you like us to know? What did you wish we asked you?

Many customers ask us about whether to use RDF graphs or LPGs (labeled property graphs). We support both, but in the long run we want to see unification in this area. We have our own ongoing project to make this happen, and that has the possibility, among other things, to make NIEM available to users of LPGs.

Stakeholder #2: Education (Use Case 14, Learning Resources)

Please tell us generally how a cross-domain ontology, with a resulting NIEM core ontology, could facilitate developing and disseminating Learning Resources for OKN Technologies

The Education Gateway's mission is to provide a coherent and centralized set of learning materials that educate internal and external stakeholders on the content, usage, and technologies that constitute and support the OKN; this includes (a) explaining domain interrelation (i.e., mappings, alignments, and interactions) within the OKN and (b) documentation on how to extend the OKN to cover new domains. Outcomes regarding contributors' focus on well-documented and rigorous interoperability will simplify these two core missions.

What processes could we undertake, or what products could the NIEM Ontology effort produce (other than the ontology and resulting knowledge graph itself), to best facilitate the Education OKUN's mission?

The Education Gateway is committed to providing learning materials tailored to different audiences,

where each will have different learning objectives and perspectives. In order to be maximally effective, Use Case D stakeholders should maintain analogous resources. For example, one user group may be an owner of data already annotated with NIEM XML, but needs instruction on migrating OWL. Another example may be a pair of OKN stakeholder groups that need to ensure their data is interoperable. A final example would be an external user group that wishes to utilize the OKN for their own purposes and thus require high-quality documentation on understanding NIEM and its use within the OKN.

What else would you like us to know? What did you wish we had asked you?

What are the key technologies you will be using to do your work? What kinds of technologies will your users need to be able to navigate and understand in order to use your resource?

Human-centered design

Group D contributors interviewed users to identify their needs and wants for a NIEM ontology. Users and their responses are listed below.

User #1: Decarceration

They will use NIEM's Service Plan construct, as well as NIEM's messaging to pass a Notification of Release.

User #2: Homelessness

We would build APIs for both the homelessness domain and for cross-agency and cross-functional collaboration. These APIs would facilitate collaboration to end homelessness and further enhance services to people facing homelessness. HUD does not feel it has authority to specify these technical artifacts, but the country badly needs HMIS APIs.

We would use the ontology to specify the serialization of RDF, to serve as a more descriptive currency for homelessness data exchange than existing XML (XSD based) and CSV formats

User #3: Space Biology

What is the value premise?

- There is no true space biology domain ontology right now
- Biological assays require an ontology to match dependent variables between databases
- Map terminology between experiments and databases: a very common problem in biological sciences
- In spaceflight, your experiment is linked to all other aspects of the flight. Mission>Payload>Hardware>Experimental parameters. All these terms/attributes need to be linked/harmonized in an ontology with a hierarchical relationship.

- Space biology is done across the world; international agencies will enable coordinating efforts and linking terms across agencies
- Space biology is a niche in biology; could use an ontology to relate to other biomedical / natural sciences research

What kinds of inquiries could we make with an ontology/KG that we couldn't make today?

- Knowledge graph machine learning based analysis between multiple datasets, experiments
- Straightforward user data submission with well-defined, harmonized terminology and relationships (this enables Open Science and data reuse/sharing)
- Limitations in federating with other databases to data mine

How would the existence of a KG based ontology enhance your mission? Would it become an integral part of your operation?

- From the perspective of data curation/databases, simplify data collection/entry
- From a user perspective, simplify data/variable/attribute selection
- KG-based ontology would drive integration of multiple data sources and data types in a way that has previously been impossible

Ethics & social implications

As an enabling and foundational use case, the NIEM Ontology will provide appropriate means to define and make inferences about aspects like privacy and confidentiality. Specific data considerations will need to be determined by experts in those particular use cases. The NIEM Ontology provides a means to instantiate those considerations.
This will allow for better inferences, better flowing of consent, better handling of privacy, and better notification of a need for human intervention.

Ethical considerations for the NIEM ontology are most relevant in the applications of the ontology to satisfy the requirements of any particular use case or open knowledge network derived in accordance with the NIEM ontology or a catalog of ontologies designed to be NIEM conformant. It is in the context of use that issues of privacy and other relevant ethical considerations arise.

In the application of knowledge graph technology to implement an open knowledge network, there will inevitably be an associated use of artificial intelligence or machine learning components. It seems reasonable therefore, to insist that the NIEM ontology or such a library would enable adherence to the ethical principles that have been or are being developed internationally for artificial intelligence.

At the very least, the NIEM ontology should not be inconsistent with the AI ethical principles and should provide for the preservation of such principles in the implementation of the ontology through metadata or overlays or some other constraint on the use of the ontology.

To the extent that this construct is useful, we can then explore the established and emerging

international work regarding ethical principles of AI for the purpose of defining how such principles could be applied to and made an essential component of the NIEM ontology.

Many international efforts are underway or have developed ethical principles for AI. A few examples include:

- World Health Organization Guidance for Ethics and Governance of AI for Health
- OECD Council on Artificial Intelligence
- Council of Europe Commissioner for Human Rights
- The European Commission Ethics Guidelines for Trustworthy AI
- Japan guidelines on the use of AI, including on research and development and utilization
- China National Governance Principles for the New Generation Artificial Intelligence
- Beijing Artificial Intelligence Principles
- Asia's first Model AI governance framework
- Singapore: international industry-led Advisory Council on the Ethical Use of AI and Data
- The African Union's High-level Panel on Emerging Technologies broad guidance on the use of AI
- U.S. Department of Defense Innovation Unit Responsible AI Guidelines

Path to success

Group D contributors identified connectivity, sustainability, and scalability as essential to the ongoing success of the NIEM ontology. They interviewed contributors to other use cases to identify possible connections between knowledge graphs.

They identified both short-term (1 to 3 years) efforts and long-term (4 to 15 years) efforts and resources needed to establish and maintain a NIEM ontology.

Leveraging NIEM by moving it from a common vocabulary to a common understanding would enable and catalyze connecting the other use cases, a horizontal use case to connect these vertical use cases. This is because knowledge graph based data integration relies centrally on creating mappings between the underlying schemas (i.e., ontologies) of the different graphs. By reusing NIEM ontology modules in other OKN efforts, this mapping creation becomes much easier.

This is crucially important for privacy and consent. Concerns about these already cross domains.

A concrete step-wise implementation plan would be as follows:

1. Define scope of ontology modeling effort by means of brief description of a number of (diverse) use cases.
2. Make a priority list of the NIEM core terms that informs the sequence of developing corresponding ontology modules.
3. Model these as drafts (using diagrams)
4. Assess drafts as to fit for previously defined scope and whether they have a favorable level of

deviation (or not) from the current NIEM model.

5. Iterate 1-4 if revisions are needed, and/or progress with the next set of terms from the priority list, until the complete list has been processed.
6. Finalize formal model (documentation, OWL etc. files, mappings to and from NIEM core).
7. Develop mappings to other models if desired.
8. Develop a detailed demonstrator use case (i.e. demonstration graph based on the NIEM ontology)

This description generally follows established ontology modeling processes (e.g. **Shimizu, Hammar, Hitzler, Modular Ontology Modeling, Semantic Web**).

In the longer term, the goal would be to make the NIEM ontology support and evolution part of the more general NIEM effort, using the established and evolving long-term mechanisms.

Other concepts & issues

Group D contributors identified additional concepts relevant to the success of this knowledge graph. They also identified additional issues to be considered. Key success factors include:

- The need to establish Niem as a standard for building exchanges.
  - This standard should apply across domains.
  - It should include a governance process that reflects stakeholder input.
- Collaboration mechanisms
- Funding

Group E: Connecting investments in nutrition security and climate

Group E explored how creating a food and climate OKN could facilitate communication between federal agencies charged with protecting nutrition security and mitigating and planning for climate change. This OKN could increase understanding among agencies to support interagency efforts to mitigate these issues.

Contributors

Group E contributors are academicians. They explored the need for connecting investments in nutrition security and climate and identified possible solutions.

Group E contributors and their organizations are listed below.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Pascal Hitzler	Kansas State University
Sudip Bhattacharjee	University of Connecticut
John MacMullen	University of Illinois Urbana-Champaign
Shashi Shekhar	University of Minnesota
Ann Stapleton	U.S. Department of Agriculture
Ilya Zaslavsky	University of California San Diego
Krzysztof Janowicz	University of California Santa Barbara
Douglas Rao	National Oceanic and Atmospheric Administration
Matthew Lange	IC Foods
Sharat Israni	University of California San Francisco
Ellie Young	Common Action

Champion

Ann Stapleton of the U.S. Department of Agriculture was the champion for Group E. As champion, Ms. Stapleton organized use-case group meetings and briefed contributors on group updates.

Serving the public good through investment & connection

Today, interagency working group communication person-to-person communication provides the only means for understanding relevant data investments across federal agencies and departments. No scalable method exists. A food and climate open knowledge network (OKN) has the potential to vastly scale up what is known about federal investments, which is key to supporting our interagency efforts and in meeting White House all-of-government priorities.

Nutrition security and climate change responses are connected in many ways at many levels of granularity. Government agencies have an understanding of what actions need to be taken across these areas, and clear executive orders making these actions a priority, but they have less information on how current investments relate to each other to build the larger federal response. This OKN would address that information gap.

The collective responses from interviews conducted with interagency working groups were used to produce the word cloud (**figure 2**) below.

Figure 2: Word cloud — Intersections between climate, nutrition security, and data

Goals & objectives

Integrate and apply existing climate, environment, food system, and nutrition/health ontologies to public agency award data to build the knowledge graph connecting climate and food system investments, their outputs and outcomes. Disseminate the results as a living, queryable graph (including gaps in knowledge) via interagency climate and nutrition security working groups.

End-users

The Connecting Investments in Nutrition Security and Climate OKN will support end-users in the federal government; that is, anyone making decisions about federal investments — such as senior agency leaders, program leaders who propose areas of investment, and strategic planners. Projects such as GAO's high-priority program inventory would be enhanced by this OKN use case. A public knowledge graph with this data would also serve general government transparency and enhanced public access purposes. It would additionally support potential nutrition and climate researchers interested in crafting proposals, programs, and projects for addressing current nutrition and climate issues.

Resources for a Nutrition Security & Climate OKN

Group E contributors identified government agencies charged with protecting nutrition security and with preparing for and mitigating the effects of climate change. They then identified resources relevant to OKN development, including datasets, knowledge graphs, ontologies and AI/ML/DL techniques.

Government Agencies

Government agencies and agency resources that could be used for Use Case E development include:

- United States Department of Agriculture National Institute of Food and Agriculture (USDA NIFA)
- USDA National Institute Food and Agriculture, Gus Schumacher Nutrition Incentive Program (GusNIP)
- USDA Agricultural Marketing Service (AMS)
- USDA Community Food Projects (CFP)
- National Science Foundation (NSF) **awards data portals** [opens in new window]
- National Institutes of Health (NIH) **nutrition awards** [opens in new window]

Ontologies
Ontologies that may be useful in OKN development include:

- The Agronomy Ontology (AGRO)
- The Chemical Entities of Biological Interest (ChEBI)
- The Farm to Fork Food Ontology (FoodOn)
- The Compositional Dietary Nutrition Ontology (CDNO)
- The Human Disease Ontology (DOID)
- The Medical Action Ontology (MAxO)
- The Health Surveillance Ontology (HSO)
- Environmental Conditions, Treatments, and Exposures Ontology (ECTO)

- The Food-Biomarker Ontology (FOBI)
- The Ontology for Nutritional Studies (ONS)
- The Ontology for Nutritional Epidemiology (ONE)
- The Process and Observation Ontology (PO2)

Knowledge graphs

Knowledge graphs include:

- WeAdapt
- KnowWhereGraph

The KnowWhereGraph already has an extensive spatial knowledge graph (over 12 billion triples) that integrates climate- and food-related data relevant to food supply chains and disaster response. This knowledge graph can possibly be leveraged. KnowWhereGraph furthermore has sophisticated support for spatial data (including integration with ArcGIS and QGIS) that can be leveraged, and this includes some integration across different granularities. The graph has already been released and software will also be released soon to the public for re-use.

Datasets

Unified aggregation of the datasets, ontology tags (and topic tags if they are generated), and viewing of the results across agencies are a must-have.

Datasets include:

- Datasets available for different geographies:
  - Agricultural produce
  - Water use
  - Land use
  - Temperature, humidity and other local information
  - Food and nutrition outlets
  - Some socioeconomic status of consumers (social determinants of health).
- Datasets not available or not available publicly:
  - All datasets are not available at the same geographic levels of granularity.
  - Detailed socio-economic and household composition information.
  - Detailed information about nutrition outlets and products available.
  - A supply chain perspective of the linkages between these elements.

Other expertise

Group E contributors identified additional areas of expertise that could increase the impact and scope of the Connecting Investments to Nutrition Security and Climate knowledge graph. They determined that data.gov staff must be contacted to gather their requirements for increasing this knowledge graph's impact and scope. They also identified the NIEM Ontology knowledge graph (see Use Case D above) as a useful resource for this effort.

Areas of expertise

Group E contributors plan to reach out to data.gov staff to determine what would be needed for them to provide this functionality.

The parallel NIEM Ontology effort (Use Case D) may also be helpful; in fact, some NIEM vocabulary may be directly usable as integration format, once ontologized. Additional expertise that could be added includes economics, systems engineering, systems dynamics, network analysis, data and record linkage, big data analysis, and machine-learning tools.

Required resources for a minimum viable product (MVP)

The following resources are needed to develop a *Minimum Viable "Product" (knowledge network):*

- **Datasets**
  A known-truth data set or data set generating simulation to check that the ontology tagging is giving the right output. This will help determine the need for topic modeling/natural language processing as an intermediate step.
- **Tools**
  Visualization tools beyond returning lists will be required. This could be **RShiny** or **Dash**, but this is a must have capability on the query platform.
- **Infrastructure**
  Ability to load datasets to data.gov and computational resources for the processing, and servers to host the query and results viewing functionality.

Unknown and Future Resource Requirements

Following concerns still exist regarding resource unknowns: Will data.gov work? Why or why not? If not, what else would be 'owned' by GAO that could be inherently interagency for hosting and serving visuals? How would long-term maintenance, operations, and refresh be ensured?

A maintenance and enhancement plan and interactive documentation would be needed so that people can help each other. An interactive "this data point is not labeled correctly" button that feeds back to modeling is also needed. This would allow models to be continuously updated and there is no risk of model drift. Continual updates include suitable user feedback regarding graph content.

End-users & stakeholders

Probable end-users include federal employees on interagency working groups, program leaders, OSTP/ policy leaders, the public, and academic policy researchers. To serve the public good, end-users would utilize this knowledge graph to find connections with other OKNs. They would use insights obtained from these to allocate investments in nutrition security and climate effectively, reducing costs and ensuring that both areas receive sufficient funding to improve nutrition among the public and to mitigate the effects of climate change.

Probable stakeholders include technical experts on OKNs, including schema/ontology specialists, data portal managers (who might need to push data to data.gov), interagency working group members who define key questions and use the information regularly, data.gov/GAO IT resource managers and technical staff.

Human-centered design

Group E contributors collected a range of example queries from interagency working group members to identify end-users' wants and needs. Some queries were very big-picture and some were very focused on specifics. Specifics are useful in building the underlying OKN structure.

The most common comment from interagency participants was some version of "This would be useful now. When can we access this tool?"

Contributors were unable to reach **data.gov** staff or similar experts during the OKN sprint.

Ethics & social implications

Group E contributors determined that this knowledge graph project would only integrate publicly available data, including data about research grants. Any supplemental data, such as inferred geotags, would need to follow specific agency guidelines. Governance would need to be managed differently if it were hosted at +data.gov versus outside of government.

Path to success

Use Case E contributors explored connectivity, sustainability, and scalability as essential to the OKN's success.

Connectivity

Group E contributors determined that some customization of technical details would be required if this were government-server-hosted and maintained, but the architecture can and should be aligned with overall general OKNs.

An overall OKN would require a careful alignment of existing suitable schemas (a.k.a. ontologies) and methods to determine if existing ontologies are adequate. The ontologies could be strongly informed (and aligned with) the NIEM ontology (**Use Case D**) that is being developed in parallel, and likewise with the innovation sprint **Group P**, which focused on natural disasters. This group's use case seems to be highly synergistic with **KnowWhereGraph**. Interconnecting with the KnowWhereGraph graph and schema could also be done naturally.

Sustainability and scalability

If this were part of **data.gov** and pulled public data from federal agencies' servers, a defined funding stream would be needed for development and maintenance. Ways to scalability and repeatedly disseminate availability and value to users and funders.

On the *data* side, sustainability is closely related to quality graph (and graph schema/ontology) design. Hence, a reasonable portion of the project effort needs to be devoted to this, and to schema-level and data-level integration with other OKN graphs. In particular, this needs larger resources early on, and much fewer resources later (if done well initially).

On the *governance* side, clear workflows need to be established involving interested stakeholders, as well as a stable governance, in order to guarantee long-term sustainability.

Scalability is multifaceted. Graph construction will require a significant effort in terms of person hours. Graphs with more than 10B triples can be hosted by good triple stores. Querying scalability will have to be looked at more closely as the use cases are more clearly described.

Other concepts and issues

Group E contributors identified additional concepts crucial to the success of an OKN as well as additional issues that should be considered.

For integration between the different OKN efforts — in order to arrive at an integrated OKN that is useful as such, it is important to harmonize schema development across the different sub-efforts. If this is not done, then only loosely integrated graphs can be used together without additional significant effort. (This is the fate of the linked data cloud, which does not have schema/ontology harmonization.)

Group F: An Open Knowledge Network for Health Communication Management

Group F contributors propose to build a knowledge graph for health communication management, as part of the Proto-OKN, that helps health communicators to better understand the landscape of online health communication and to create, discover, and share effective messages for promoting health behaviors and mitigating health misinformation.

Contributors

Contributors to Use Case F are listed below, along with their organization.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Nariman Ammar	The University of Tennessee Health Science Center
Irma Arispe	The Centers for Disease Control and Prevention
Erin Carlson	The University of Texas at Arlington
Chengkai Li	The University of Texas at Arlington
Brian Handspicker	Open City Labs
Florence Hudson	Columbia University
Dagny Olivares	Center for Disease Control and Prevention
Katerine Osatuke	Department of Veterans Affairs
Mark Schildhauer	University of California, Santa Barbara
Ilya Zaslavsky	University of California San Diego

Champion

Chengkai Li, Professor, Associate Chair of the Department of Computer Science and Engineering, and Director of the Innovative Data Intelligence Research Laboratory (IDIR) at the University of Texas at Arlington, was champion for Group F. As champion, Prof. Li organized use-case group meetings and briefed contributors on group updates.

Goals & objectives

An OKN for Health Communication Management would meet the following goals and objectives:

- This will improve public health by promoting health behaviors and mitigating health misinformation, which is particularly crucial for vulnerable populations such as older adults, low-income families, and minority communities.
- Why OKN: It promotes accessibility and interconnectivity, as it allows for quick bootstrap, expansion by more contributors, and integration with other datasets
- To facilitate the use of this OKN, we will build a public health communication management platform using the OKN. We will make the OKN accessible so that app developers can use it to create their own tools.
- The main users are health communicators, (e.g., public health offices, health practitioners, media, and scientists), as well as citizens of diverse backgrounds. This OKN will become a "go-to" site for a broad range of stakeholders to (1) seek the latest understanding of the prevalence and receptions of online health narratives; (2) to discover effective intervention messages produced by credible, reliable sources, and (3) to request, publish, and share such interventions.

Particular focus will be on the contextualization of intervention messages based on message topics and target audience, in order to make health communication comprehensible and trustworthy to a broad range of users.

The secondary users can come from other sectors. Greater understanding of health communication and interventions could help develop approaches for other domains, (e.g., transportation and safety information dissemination and interventions).

Resources for the Health Communication OKN

Group F contributors identified the following components of this OKN:

- Online health narratives, including misinformation, from social media (e.g., Twitter), discussion forums, web pages, and so on.
  - For example, "Study indicates that vaccinated children have more health issues than unvaccinated children …"; "FDA approved Ivermectin for human use in Feb 1996."
  - This is the Minimum Viable Product (MVP) of the project.
- Sources of the narratives: Influential organizations and individuals that started and spread the narratives
- Responses to the narratives (e.g., likes, shares, sentiment, and stance)

- Truthfulness assessment (e.g., fact-checks) of the narratives from public health departments, hospitals, health professionals, reporters, and researchers
- Intervention resources: messages promoting health behaviors and debunking misinformation
- Integration with health and medical knowledge bases (e.g., WebMD)
- Reference documents: Research literature, media resource links, and so on

Existing data/resources

Group F identified the following existing data resources that could be used in this knowledge graph's development:

- Social media APIs, social listening platforms
- Claim spotting tool **ClaimBuster**
- Tweet monitoring tool **ClaimPortal**
- Health-related FAQs from websites such as CDC, Mayo Clinic, WebMD, etc.
- Health-related fact-checks from Metafact, PolitiFact, and so on

Data sources to be created or acquired

Use Case F contributors identified the following data/resources to be created or acquired:

- Lists of public health departments, hospitals, health professionals, reporters, researchers
- Ongoing development of unique claim detection, truthfulness stance detection, claim mapping algorithms and APIs
- Integration with health and medical knowledge bases (e.g., WebMD)
- Integration with publication database (e.g., PubMed)

Other expertise

Group F contributors identified the following additional areas of expertise that could increase the impact and scope of the Health Communication Management knowledge graph.

- Public health communication
- Health psychology

Health Communication OKN end-users and stakeholders

The Health Communication OKN serves the public good by disseminating accurate and trustworthy communications about health issues that affect citizens and residents nationwide. This knowledge graph would quickly acquire and analyze data pertaining to widespread and potentially dangerous health issues and use it to inform and advise the public on health-safety measures.

Group F contributors identified the following primary end-users and stakeholders of this knowledge graph:

End-users

The primary end-users of this knowledge graph will be health communicators, including:

- Public health offices
- Public health systems (e.g., hospitals)
- Health care practitioners (e.g., doctors and nurses)
- Organizational communication offices
- Health reporters
- Health researchers and educators

This knowledge graph would help health communicators:

- Better understand the landscape of online health communication
- Discover, create, and share effective messages to promote health behaviors and mitigate health misinformation

It would also provide end-users with access to advanced analytics for assessing the efficacy of communications. Advanced analytics capability includes:

- Similarity search: given an online health narrative, find similar narratives that exist
- Importance/centrality: measure how influential a narrative is
- Source detection: identify who started the narrative
- Diffusion patterns: how did the narrative spread and evolve
- Prediction: how viral will the narrative become
- Credibility assessment: summarize relative credibility of different pieces of information; let people compare evidential basis for all types of information and their explanatory chains

Stakeholders

Stakeholders essential to the development and deployment of this knowledge graph include:

- Social media managers, content strategists, coordinators, and producers
- Developers of social listening tools
- Users of publication servers
- Users of health/medical knowledge databases
- Data science educators

Human-centered design

Group F contributors interviewed potential end-users and stakeholders to obtain their requirements for the user interface of the Health Communication Management OKN. Briefly, the OKN UI should be easy to navigate for all users of the OKN. The UI should provide easy access to information resources specific to this OKN and should display information in easily understood graphs and text. Potential end-users and stakeholders and their names and organizations are listed below.

Potential end-users

- Katerine Osatuke, Research Director, Department of Veterans Affairs
- Tiffany Kindratt, Assistant Professor, Public Health Program, University of Texas at Arlington
- Erin Carlson, Director, Graduate Public Health Programs, University of Texas at Arlington

Potential stakeholders

- Irma Arispe, Director, Division of Analysis and Epidemiology, National Center for Health Statistics, CDC
- Dagny Olivares, Associate Director for Communication, National Center for Health Statistics, CDC
- Florence Hudson (representing Group N on OKN Education)

Ethics & social implications

Use Case F contributors identified the following ethical implications of a Health Communication Management knowledge graph/OKN:

- No personal data or health records can be collected or publicized.
- Governance for this knowledge graph must be established. Governance must cover the issues:
    - Which social media accounts are to be included and how should they be categorized/labeled? (Should every doctor be trusted?)
    - Which organization's truthfulness assessment of narratives should be included? For example, whose fact-check is a true fact-check?
    - Accommodation of different views
    - Resolution of conflicts
    - Up-to-date data Respect for social media users' privacy: Focus on narratives instead of individuals.
    - Consent framework for enabling third-party app development on top of the data (e.g., examples from **digi.me**)

Path to success

Group F contributors recommend activities and resources for short-term (1-3 years) and long-term (4-15 years) sustainability and scalability.

Short-term activities

In the short term, a few activities are crucial in order to sustain the dataset and use case:

1. Keep data collection, cleaning, and injection tools running continuously in order to ensure data is up to date.
2. Maintain data collection, cleaning, and injection tools so that data is successfully collected even when data sources undergo changes.
3. Keep building health communication management apps and tools in order to help users utilize the OKN.

4. Expand the data content of the OKN to include information beyond health narratives (e.g., information about narrative sources, responses, truthfulness assessment, intervention resources, and so on).

Long-term activities

In the longer term, the aforementioned activities for the short-term remain crucial for sustainability and scalability. On top of that, growing a community of users and stakeholders will be key. Furthermore, computing and storage resources are important for ensuring the continuity of the project.

Group G: Enabling Community Climate Action
And Local-Federal Intelligence Transmission

Use Case G was selected for development based on Biden Administration executive orders to combat climate change and the need to support climate-change initiatives at the local level.

Contributors

Contributors, listed below, are volunteers representing key stakeholders from academia, government, and industry. Stakeholders have direct experience with or interest in this issue.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Stan Ahalt	University of North Carolina at Chapel Hill
Ashley Atkins	University of California Berkeley
Brian Blanton	University of North Carolina at Chapel Hill
Chris Bizon	University of North Carolina at Chapel Hill
Johannes Friedrich	World Resources Institute
Matt Gee	BrightHive
Bandana Kar	Oak Ridge National Laboratory
SJ Klein	Harvard University
Chris Lenhardt	University of North Carolina at Chapel Hill
Ryan McGranaghan	National Aeronautics and Space Administration
Douglas Rao	National Oceanic and Atmospheric Administration
Lilit Yeghiazarian	University of Cincinnati
Ilya Zaslavsk	University of California San Diego

Champion

Ellie Young, founder of Common Action, was champion for Use Case G. As champion, Ms. Young organized use-case group meetings and briefed contributors on group updates.

Supporting local community climate-change initiatives

Although mobilizing climate change action is an urgent national priority, mitigation efforts rely primarily on local implementation efforts. Under the Biden Administration, climate change has become an official national priority.

The Build Back Better Act and various executive orders, such as the Executive Order to Strengthen America's Forests, Boost Wildfire Resilience, and Combat Global Deforestation, call for ambitious transformations across a range of climate-related systems, including energy, water, agriculture, the built environment, and industrial production. While achieving these targets requires broad efforts throughout the nation, climate change mitigation and impact occur primarily in local contexts within cities and communities.

Currently, the capacity of local cities to mitigate climate emissions and implement adaptation and resilience initiatives is wholly dependent on locally available expertise and resources. To augment local cities' ability to identify immediate climate initiatives, set necessary goals for longer-term mitigation and resilience sufficiency, and develop capacity to reach these targets, we propose to develop a national knowledge base to describe high-level climate action targets and related resources, shaped to fit local city operational environments. This undertaking requires an assessment of the current state of practice across national and international climate plans and implementation frameworks, as well as an assessment of goals, capacity, gaps, and barriers generally common across cities.

Goals & objectives

Goals and objectives for Use Case G are to support the efforts of U.S. cities to fight climate change. Barriers to these efforts are discussed next.

Barriers for cities to reach climate objectives

Cities are the fundamental locus point of mitigation, adaptation, and resilience efforts. The most recent **IPCC report** recognizes cities at large as a major source of carbon emissions, citing consumption, electricity, and transit as primary drivers. Industrial carbon emissions occur at sites located within cities, and are operated to produce goods that are ultimately consumed within cities.

The systems that adaptation and resilience efforts are designed to protect — water, electricity, and so on — form the foundation of city life and are typically managed by municipalities. As climate change continues, overall social well-being will be determined through feedback loops between changing environmental conditions, management response, and resulting resilience of social and environmental systems.

Role of municipalities in climate-change mitigation

While cities' contributions to climate change are fueled by both public and private-sector activities, municipalities play a central role in guiding action towards mitigation and ensuring foundational stability. Broadly, their responsibilities include:

- Incentivizing the reduction of carbon emissions from energy use, through transition to renewable energy, and reduction of building and transit energy consumption
- Reducing energy usage and anticipating changing future contexts in planning infrastructure and planning projects
- Responding to disruption events, such as natural disasters, which includes predicting disaster events, anticipating and preparing for impact, meeting complex social needs during and immediately after a disaster event, repairing infrastructure, and facilitating the rehabilitation of disrupted economic and social activities
- Providing continuous utilities (e.g., energy, water, housing) within changing environmental contexts
- Continuing to meet and develop primary objectives for human health, equity, and safety, etc.

These tasks represent a significant extension of traditional city management. Executing these activities sufficiently will require the integration of climate mitigation and resilience goals throughout city operations, which in turn requires the rapid development of new bases of knowledge and capability for decision makers, such as:

- Awareness of need for action on mitigation, adaptation, and resilience goals
- Knowledge of action options and core strategic, tactical considerations
- Capacity to engage with stakeholders to identify topics of concern, connect to climate objectives, and build support for necessary undertakings
- Awareness of systems in order to meet technical objectives, such as prediction of water disruptions, etc.
- Awareness of the complex inter-dynamics between multiple systems which influence core social objectives, such as population health, across dynamic spatial and temporal horizons
- Interpretation of changing climate impacts and implications for future planning activities

These activities require a tremendous amount of high-quality, technical information. However, tools to guide city mitigation and emissions planning consist primarily of high-level materials such as **UNFCC's Race to Zero** campaign materials. Such materials lay out aspirational directions, but do not provide strategic and tactical insights required for day-to-day decision making, and are ultimately insufficient to drive the speed and scale of change required.

For example, in a recent conversation with Erick Shambarger, Director of Environmental Sustainability for the City of Milwaukee, lack of access to basic data to inform decision-making was cited as an "infuriating" barrier to his department's ability to undertake climate change action. Mr. Shambarger additionally described other technical knowledge challenges:

- Laborious integration of federal datasets to identify carbon emissions for the entire United States —

in order to define targets for Milwaukee
- Challenge of navigating between highly complex and specific programmatic funding schemes across disparate local, state, and federal agencies — echoed by other departments in the same city.

Knowledge graphs solution for multiple information challenges

Information discovery challenges are well recognized in the fields of computer science/knowledge management, which has developed a tool to address them: knowledge graphs.

Knowledge graphs (KGs) are a "novel paradigm for the representation, retrieval, and integration of data from highly heterogeneous sources." Knowledge graphs can be developed to address multiple information challenges such as

- Integrating diverse data sources to define the scope of climate change action required nationally
- Comparing cities' existing climate-strategy plans to requirements, as well as comparing them to each other
- Analyzing available funding, datasets, and programs to define various resources and gaps
- Collecting resources about various tactical approaches, such as those provided by various nonprofits

Knowledge graphs identify knowledge gaps and inform research initiatives

By providing a centralized knowledge infrastructure, knowledge graphs can also be used to accelerate the identification of knowledge gaps and inform subsequent research initiatives.

Specialized knowledge graphs may also be developed to centralize large quantities of knowledge necessary to solve specific, yet complex problems, such as identifying real-time flood risks (**UF OKN**); integrating layered geographic data for environmental decision-makers such as disaster responders, or insurance agencies (**KnowWhereGraph**), and balancing the energy load and resilience of electricity grids. Multiple knowledge graphs may also be connected into a framework, through an open knowledge network, to address a broad range of sophisticated challenges.

Knowledge Graphs produce indirect social benefits

Building a knowledge graph also produces several indirect social benefits. Firstly, developing a requirement requires communities to agree on shared standards, which can produce a foundation for increased coherence and coordination thereafter between groups. The process of developing such standards can also serve as a catalyst for stakeholders to agree to share proprietary data that would not otherwise be made available.

For example, in other exploratory conversations conducted by our group, energy utility companies expressed interest in establishing access to proprietary datasets–as long as the level of access was coordinated across their peers as well. A knowledge graph project provides exactly the context to initiate this time of collaborative discussion.

Knowledge graphs serve as nexus point for community best practices development

Lastly, as a central base for stakeholders to gather regularly to obtain valuable knowledge, knowledge graphs also have the potential to form a nexus point around which a community of practice can develop. Such a community may be able to share best practices, coordinate with distributed members of the network, and ultimately be more open to responding to the priorities of other stakeholders.

Ultimately, such a platform may address the central barriers of climate change action: lack of focused direction and coordination capacity across disparate spatio-temporal processes and the stakeholders associated with those processes, including communications about needs and resources.

Group G contributors hypothesize that local municipalities' climate action decision-makers face a broad range of knowledge and data access gaps. They anticipate that these will conform to common knowledge management challenges, which can be addressed through knowledge graphs to facilitate baseline capacity for climate mitigation, adaptation and resilience for cities across the nation.

Therefore, they propose to undertake research to identify common challenges faced by climate action decision makers in local cities, in order to evaluate how knowledge graphs may be able to support stakeholders to reach their goals, and inform subsequent design of any such system. And how they can be addressed with knowledge tools, or from focus areas for new emphasis.

Policy-making end-users benefit from data that Informs decision-making

Professional decision-makers, such as managers of community infrastructure systems, can benefit from data revealing complex connections between converging systems, such as the agriculture-energy-water nexus. Municipalities and policymakers can also benefit from data that:

- Reveals the interrelationship between social, economic, and environmental factors, such as:
  - Multi-causal pathways and differentials between community health, environmental resources, and socio-economic status
  - Connections between animal behavior and weather, and interplay with human activity
  - Differential policy outcomes or impact experiences by geography, time span, neighborhood, identity factors, and so on.

Further mobilization can be achieved by enabling community stakeholders, such as residents, neighborhoods, and organizations, to identify and evaluate proactive strategies to achieve community climate action targets, respond to evolving local conditions, and aim for high-level targets such as those outlined in the **IPCC Climate Change 2022: Impacts, Adaptation and Vulnerability report**. For instance, residents and community leaders can explore the impact of individual and neighborhood-level projects, such as:

- The cost savings and energy reductions realized by installing solar panels on 40% of roofs in the community
- Planting native species in greenspaces
- Activating community members to organize as the first line of response during disruption events

Both formal and informal management strategies can be supported with a bidirectional communications protocol, enabling community members to provide feedback on the accuracy and usefulness of datasets to data stewards; share questions with researchers and related experts; and provide real-time updates to agencies about changing local conditions.

Resources

Group G contributors identified the following resources for development of this knowledge graph/OKN:

Foundation layer

The foundation layer places community climate change and related knowledge graphs in a general to detailed order. The foundation layer includes:

- KnowWhereGraph as framework for locating geographical features
- Starting framework: resilience map to layer resources such as Transition Network as second foundation layer. The resilience map allows defined data sets to be layered atop the foundation in the order of their relationship to geographical features.

Climate data

Climate data make up the next layer of this OKN. Climate data includes:

- NOAA Climate at a Glance provides mapping at multiple scales of the climate
- NOAA Billion-Dollar Weather and Climate Disaster product provides mapping of disaster and risks associated with weather and climate hazards.
- Climate Explorer provides projected climate change for the U.S. for resilience planning for local stakeholders.
- Urban Flooding Open Knowledge Network provides specific information relevant to flooding related knowledge.
- NASA Prediction of Worldwide Energy Resources (POWER) provides solar and meteorological data to support renewable energy, building energy efficiency, and agriculture needs.
- NOAA State Climate Summaries provide biennial updates of climate information for each state.
- Species Interaction Dataset Index provides access to existing datasets that make claims on how organisms interact with each other and their environment. (biotic/abiotic interactions like pathogens/diseases, parasites, prey, symbionts). For specific projects, see:
  - Parasite Tracker
  - Big Bee
  - Global Biotonic Interactions: Covid 19
- Open Traits Network provides access to traits (or properties) of organisms (e.g., flowering time, body size, body weight)
- Global Biodiversity Information Facility (GBIF) provides access to primarily >1B species occurrence (where/when) records.
- Integrated Digitized Biocollections (iDigBio) provides access to >100M digitized specimen records of Natural History Collections in the U.S..

- **OBO Foundry / OBO Relations Ontology** and friends — a semantic "glue" that helps to integrate and align existing knowledge.
- Taxonomic information systems that aggregate and organize definitions/publications of species classifications (e.g. **Catalogue of Life**, **NCBI Taxonomy**)
- Publishers of scientific works in an open and integrated fashion (e.g., **PENSOFT**) See, for example, **Biological dark data in times of viral spillover | ASU News** and Upham, N.S. et al., 2021. **Liberating host–virus knowledge from biological dark data**. The Lancet Planetary Health.
- Climate action plans: can be organized into an information system showing how climate challenges are addressed in different parts of the country, how comprehensive they are, and if there are gaps that OKN can fill. The **San Diego climate action portal** provides a good example. Parsing climate action plans would also help formulate questions that the OKN must answer, and engage local planners beyond climate resilience experts.
- **Data Translator API**
- **DataCommons**
- **EarthCube**

Community input

Community input makes up the next layer. This layer is made up of questions sourced from the community themselves. These include free text and questions gathered from local Google searches through the engagement process.

Ontologies

Ontologies are layered atop the Community Input layer. Existing ontologies to be used for this OKN include:

- **PPoP Ontology**
- **DNA Ontology**

Ontologies that are necessary to develop include:

- Climate action ontology
- Domain ontologies for each of these data areas

Other expertise

Group G contributors identified areas of expertise that could increase the impact and scope of the Enabling Community Climate Action and Local-Federal Transmission knowledge graph/OKN. Key among these are community engagement experts and representatives of federal agencies, such as the NOAA, U.S.GS, and NIEHS.

They recommended that a resource committee consisting of representatives of various agencies and domains be organized to provide agency- and domain-specific expertise. They also identified the following issues that these community experts and agency representatives would need to address.

These include:

- As the proposed OKN is heavily leaning towards engaging with the community, it is crucial to have community engagement experts to facilitate the engagement and "translate" information across scientists and community members.
- We should also think about how to assess the impact and value of the proposed OKN. Current assessment framework including value of information (VOI) can be used. But we should also think about other participatory assessment tools.
- Ask NSF to connect us with climate resilience experts at federal agencies (NOAA, U.S.GS, NIEHS), to see how they would benefit from the OKN, and help formulate questions
- A resource committee of folks spanning agencies/domains, to provide guidance as needed

End-users & stakeholders

In many communities across the United States, climate change will critically impact water supply systems, through disruption events such as floods and broader water quality and quantity pressures.

Water is influenced by virtually all socio-environmental activities, including those that determine environmental and social resilience to climate impacts. These include:

- Discharges from industrial and municipal sources; urban, agricultural, or other forms of polluted runoff; depleted or contaminated ground water
- Landscape modification; changes in flow
- Overharvesting of fish and other organisms; introduction of exotic species; bioaccumulation of toxins
- Deposition or recycling of pollutants between air, land and water.

Thus, water management provides a focus point through which to address 1) critical infrastructure, 2) disaster impact, 3) landscape resilience 4) community action.

Managing at the watershed level, which is advocated by the EPA and U.S.DA, additionally requires 5) coordination between multiple stakeholders, which is currently organized through email and phone communication between parties who connect only intermittently.

Stakeholders

- Group G contributors interviewed Kevin Shafer, Executive Director of the Milwaukee Metropolitan Sewerage District (MMSD), to identify MMSD's wants and needs for this OKN. Mr. Shafer's responses centered on the MMSD's need to expand and enhance their climate-focused management practices. Key among these are: Working with farmers to reduce point-source pollution from agricultural runoff
- Coordinating flood response with other utilities
- Informing land-use planning

MMSD and other water management stakeholders identified these additional needs:

- Integrated model of environmental and city landscape dynamics, including mapping of soil, flood

plain, and existing land use (supplied in part by the **KnowWhereGraph** OKN)
- Real-time flood alerts (supplied by the **Urban Flood Open Knowledge Network**)
- UI support for cross-stakeholder communications and data-sharing

End-users

Use Case G contributors identified end-users of the OKN and their wants and needs. End-users include:

- Community residents
- Farmers
- Academics
- Small businesses

They then interviewed representatives of these end-user groups. All are from the Milwaukee, Wisconsin, area. They include:

- Dean Amhaus, President & CEO, The Water Council
- Kevin Muhs, Executive Director, Southeastern Wisconsin Regional Planning Commission
- Kevin Shafer, Executive Director, Milwaukee Metropolitan Sewerage District
- Erick Shambarger, Environmental Sustainability Director, City of Milwaukee
- Rebecca Klaper, Vice Dean and Professor, School of Freshwater Sciences and Director, Great Lakes Genomics Center, University of Wisconsin-Milwaukee
- Sarah Bregnant, Northwest Side Community Development Corporation

Community residents

Community residents want projects that are relevant to local concerns. Their primary concern about climate change is flooding impact. They also said that support for local and urban agriculture is an important concern. Community residents said that the user interface (UI) for this OKN should include features to help them understand and prepare for disruption-related risks so as to support decision-making in future projects.

Farmers

Because agricultural data is highly political, this end-user group wants trust and privacy features built into the OKN in order to participate and share data with it.
Use Case G contributors identified the need for an OKN feature that would meet farmers' requirements. Features would include:

- Incorporating private and designated security settings in the OKN
- Empowering local actors (such as MMSD) to build trust with farmers, using the OKN as support.

Academics

Rebecca Klaper, Vice Dean and Professor, School of Freshwater Sciences, was interviewed to obtain this end-user group's requirements.

Professor Klaper said that academics want to contribute data. She said that the OKN should allow focus on studying regional variability to better predict huge fluctuations in water levels, which influence agricultural output and the decisions of the MMSD and other decision makers charged with water management.

She said it should provide information to highlight water-related problem areas in the state. It should also make this information readily accessible to state policymakers and show its link to economic implications. This would enable sound decision-making.

Group G contributors identified the need for an OKN feature that would meet farmers' requirements. Features would include:

- Integration of federal data with local data to support academics in their efforts to:
- Research factors contributing to climate change, and
- Identify possible solutions
- OKN capability to include data contributed by academics.

Small businesses

Dean Amhaus, president of **The Water Council**, a nonprofit dedicated to solving critical water challenges, was interviewed to obtain small business requirements for this OKN. Mr. Amhaus said that small businesses lack guidance on how to reduce their water consumption and want the OKN to provide it. Use Case G contributors identified the need for an OKN feature that would meet small business requirements. Features would include:

- Access to guidance resources and expertise relevant to the water-related needs and wants of small companies.
- Capability for small businesses to evaluate the impact and cost savings of various activities.

Ethics & social implications

Group G contributors identified potential ethical implications of an OKN for enabling community climate change action. The ethical implications of an Enabling Community Climate Action and Local-Federal Intelligence Transmission OKN are detailed on this page.

Group G contributors recognized that the Community Climate Action OKN would have ethical implications throughout the entire project lifecycle. Implications are particularly relevant to the context of climate. That context is notable for the number of stakeholders, differing goals, and the political context of data.

An end-to-end approach must be taken to identify ethical decision and operations points throughout the project lifecycle, as shown in **figure 3**.

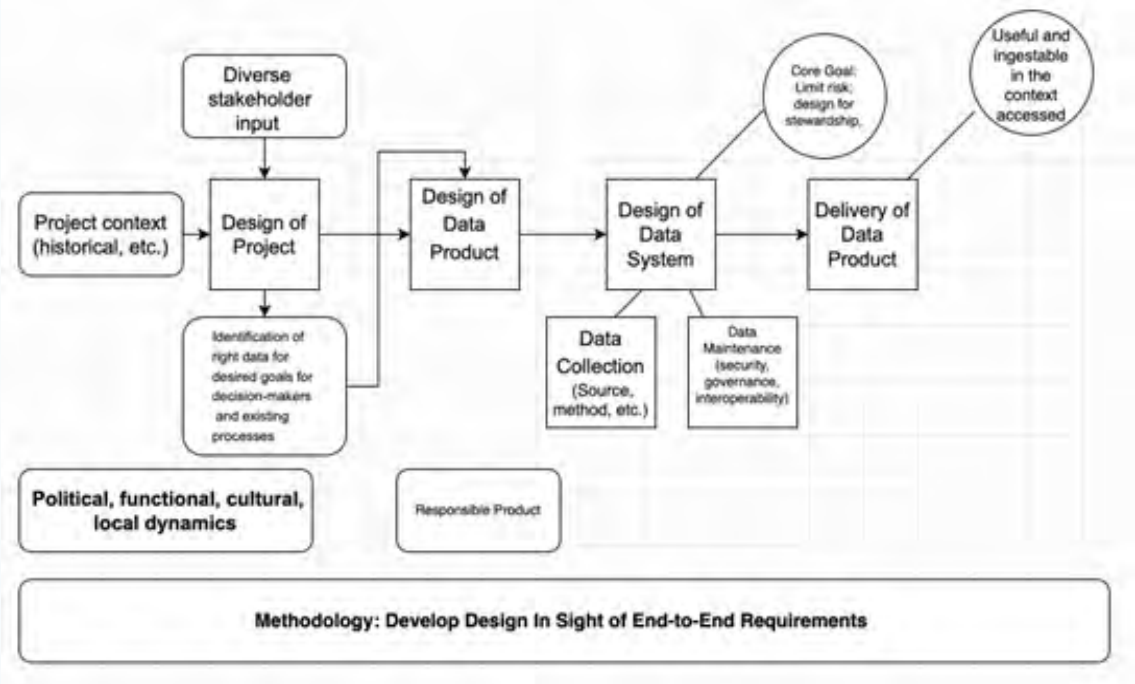


Figure 3: Methodology: Develop Design in Sight of End-to-End Requirements

The end-to-end approach includes the following decision and operations points throughout the project lifecycle:

- **Design of Project:** Design system to connect to existing community priorities; tailor the project design to a stakeholder-led design process
- **Data Product:** Designing functionality to receive, input, process data responsive to needs of stakeholders in situ
- **Data System:** Designed and maintained to ensure security, privacy, access; identify treatment methodologies for particular
- **Data Collection:** Strategy of type of data to collect and methods to reduce risk (e.g., avoid PII).
- **Data Protocols:** Identify for each type of data, including:
  - Intellectual property
  - Confidential data
  - Human level data (census tract): construct use cases to avoid PII
  - Integration/interoperability with other systems — ethical considerations of exchange
  - Combination of public and private data
  - Access to private data (private agricultural data)

Path to success

Group G contributors identified connectivity as the essential factor in achieving success.

Connectivity

Group G contributors recommend connecting two use cases to produce one instantiation of the Connecting Investments in Nutrition Security and Climate Transmission OKN. The two use cases are:

- Use Case L — OKN for Energy Management
- Use Case P — Compounding Disasters / Events (a.k.a. Wicked Problems Institute)

This instantiation would provide for energy resilience by connecting:

- Community action with grid management
- Delivery of services to community members
- Community-based disaster preparation and response efforts

Other concepts & issues

Additional concepts crucial to the success of this OKN include governance, human leadership and a community of practice, and systems operations.
Governance

Group G contributors identified the following components of effective governance for this OKN:

- Two structural layers
  - Collective stakeholder management (internal to use case)
  - Decentralized communities (horizontal across OKN)

- Four themes
  - Data: How do we ensure security and privacy of data, and dynamic access controls?
  - Control: What decisions are possible to make and how are they made?
  - Knowledge management: capturing and managing information workflows
  - Operational function: the technical delivery of the use case

OKN human leadership and a community of practice

Human leadership must be tapped and a community of practice developed to launch:

- Data policy: Guidelines about how the OKN can be accessed, securing safe, functional environments while also ensuring barriers of access are not prohibitive
  - How users roles can shift/change within the structure dynamically
  - Clarity — high-level view of what is occurring within a use case
  - How communities are going to engage with each other, external
  - Constraints around the distribution of governance controls within a community use case
  - Protocols for data, knowledge, operations management

- Problem definition
  - Who decides what problem to work on? Who can work on what?
  - How is the work performed?
  - Day-to-day and long-term planning
- Entry into the system
  - User types, access capabilities
- Technical tool development
  - Development of the materials required for the use case (UIs, etc.)
  - Support protocols (knowledge management actions, etc.)
  - Management layer of the use case

Operations of the system

Group G contributors determined how the system should operate for end users and administrators.

For users

The system must provide guidance to users on how to participate within it. Guidance must include information about- options for data collection, aggregation, sharing, and communication.

For administrators

The system must allow administrators to submit issues, such as a policy for adjudication, an action to be taken, or an appeals process.

Group H: Homelessness OKN for Cities

Use Case H was selected for development to address the crucial need to combat homelessness in the United States. The U.S. Department of Housing and Urban Development (HUD) says in a **2022 press release** of its 2021 annual report that homelessness for chronically homeless individuals increased by 20 percent between 2020 and 2021.

Group H contributors identified issues contributing to homelessness, including:

- The aggregate costs of letting homelessness continue, and
- The amortized costs of remediating and preventing homelessness

Possible solutions enabled through Homelessness OKN for Cities knowledge graphs would track homelessness in real time, identify available homelessness programs and services by city and region, and identify gaps that could be filled.

Contributors

Contributors, listed below, are volunteers representing key stakeholders from academia, government, and industry. Stakeholders have direct experience with or interest in this issue.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Jan Aasman	Franz Inc., Scalable Graph Technologies
Ying Ding	University of Texas at Austin
Tim Goliver	U.S. Department of Veterans Affairs
Dina Hooshyar	U.S. Department of Veterans Affairs
Sam Klein	Massachusetts Institute of Technology
Sarah Schwettman	Massachusetts Institute of Technology
Eric Jahn	Alexandria Consulting
Sonia Kim	Alexandria Consulting

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
William Snow	U.S. Department of Housing and Urban Development
Raj Lanka	Shelter.App — Homeless Resources

Champion

Sam Klein of the Knowledge Futures Group and Ying Ding of the School of Information, University of Texas at Austin, were champions for Use Case H. As champions, they organized use-case group meetings and briefed contributors on group updates.

Preventing future homelessness will serve the public good

Use Case H, Homelessness Knowledge Network for Cities, will give city and community leaders real-time data to track homelessness, find available housing, calculate shelter and social service usage, and find funding for homelessness programs.

These capabilities will help city and community leaders house all citizens needing housing assistance and prevent future homelessness. They will provide visibility into the health of cities and will help manage dynamic flows for maintaining equilibrium.

This OKN has two parts:

- A meta-knowledge graph of how various cities recognize and respond to homelessness
- A knowledge graph per city or region, of available programs and services (with real-time data) and with knowledge of individual homeless people

The OKN will be used:

- By city planners connecting federal, state, city programs, increasing coordination and reducing duplication
- By housing and shelter providers, community organizers, and families to identify current needs and increase visibility
- By hospitals and mental health clinics, employment agencies, food security programs, and other adjacent networks
- By researchers, economists, and policy makers forecasting future needs and designing for the future

Goals & objectives

A Homelessness Knowledge Network for Cities OKN would meet the following goals and objectives:

- Increase understanding and end homelessness, city by city
- Build the OKN with data from selected cities and federal, state, city, and community programs
- Identify primary audiences — homeless organizations, city residents, charities, and city planners
- Create dashboards that display primary services in each city and include:
  - Dashboard access to graphs that show:
    - Each city's problems and solutions
    - Related patterns and models from other cities
- Link graphs to display patterns and models generated from the databases of other cities. Graph parameters would require that patterns and models represent cities of a similar size and with similar homelessness issues. Patterns and models must display the successful outcomes of strategies employed by these cities to reduce homelessness.

Resources

Group H contributors identified the following resources for development of this knowledge graph/OKN: HUD data sources, HUD MAPs, HUD health data, and HUD system-performance measures; Veterans Affairs and other government agency data, government agency dashboards, VBA contacts, policy maps, partner OKNs, and counties. Each of these resources is detailed below.

HUD data sources

HUD data: Comprises all data collected at the aggregate level. HUD data is gathered from the local (Hybrid Management Sublayer) (HMS) of the Continuum of Care (COC) Program.

HUD exchange

The **HUD Exchange** is an online platform for providing program information, guidance, services, and tools to HUD's community partners, including state and local governments, nonprofit organizations, Continuums of Care (CoCs), Public Housing Authorities (PHAs), tribes, and partners of these organizations. Although it provides access to a number of useful resources, access may be impeded for the following reasons:

- Service matching is hard to use; tracking funding for each layer of service provision may be a possible use.
- Requires research to match data with other data sources, such as:
  - **UPenn** — **AISP (Actionable Intelligence for Social Policy)**
  - **Pittsburgh** — **Allegheny County Analytics**
  - **NYC** — **Homeless & Poverty in New York City | The Bowery Mission**
  - Los Angeles – **Open Source Homeless Initiative** specialized data and city networks
- **Point-in-Time Counts** (PIT) counts of persons, households, and household types (family, adults-only, children-only)
- Methods using Homelessness Management Information System (HMIS) sheltered data can impede implementation.

Sheltered data must be hidden within the OKN, while it must still connect with other OKN data to provide more robust answers to user queries. The fact that 88 percent of COCs rely on sheltered data obtained from the HMIS necessitates more extensive programming than would be required otherwise. That results in more programming staff hours and greater expense.

Unsheltered data is hardest to get; perhaps most relevant: this sub community would benefit the most from efforts to house everyone.

Housing Inventory Count (HIC): yields a huge amount of data, once a year. Data includes the numbers of available dwellings and their occupancy type. Except for domestic violence shelters, most listings include addresses and zip codes. If available housing is not tenant-based:

- Associating the cost of building and maintenance is difficult Funding mechanisms are available
- Focus is on the night of occupancy to prevent exceeding capacity under emergency conditions
- Supportive Services for Veteran Families (SSVF) provides prevention-housing but not a housing inventory. SSVF data identifies units rented for people covered under Rapid Rehousing Works of the nonprofit **National Alliance to End Homelessness**,
- Runaway and Youth Homelessness (RHY) programs
- Sources of funding (e.g., local, state, federal, and nonprofit) for homelessness programs

HUD System Performance Measures

These include:

- Exits to permanent housing, length-of-stay in temporary shelter, returns to homelessness from temporary shelter
- Project STAY (Services to Assist Youth) of the National Prevention Information Network, Centers for Disease Control and Prevention
- Funding effectiveness per HUD program
- Outcome effectiveness of HUD's Rural Gateway clearinghouse, which provides technical assistance, training workshops, and peer learning and resources sharing to support rural and urban housing and economic development

HUD MAPS (Multifamily Accelerated Processing) Guide

This resource identifies the locations of suitable housing. Caveats for use include:

- Unreliable boundaries
- An HUD geocode may not have a perfect FIPS (NIST Federal Information Processing Standards) match
- Required user account for access to COC **shapefiles**, which include Point-in-Time and (PIT) and Hybrid Integrated Circuit (HIC) data pertaining to community planning and development.
- Underdevelopment
  - Possible alignment with Federal Information Processing Standards (FIPS) should be explored
  - Inadequate needs assessments of homeless populations

- Communication needed between the HUD Office of Community Planning and Development (CPD) and other stakeholders to agree upon and propose legislative change

HUD health data

This CDC data identifies pre-Covid tuberculosis infection rates. Data could be used to identify the percentage of homeless persons who became homeless after becoming ill.

Service Ratio, HMIS, Prison, and Healthcare Data

These datasets provide an overview of the interrelationship between homelessness rates, street outreach, and the housing history of prison inmates. They include:

- Local ratios of service numbers to unsheltered numbers to determine needed modifications or adjustments.
- Local HMIS data, which should be to deduplicated to prevent redundancy in street outreach data
- Prison data, which includes information about the housing status of prison inmates prior to incarceration and is obtained from annual prison reports and not from aggregate national reports
- Health data obtained from entry and exit interviews with residents of temporary shelter/housing for the homeless. Interview data includes:
  - Healthcare enrollment status
  - Disability status
  - Benefits enrollment or eligibility
  - Medicaid/ Social Security Disability Insurance (SSDI)

- Updates to the annual reporting requirements of contributing data sources/knowledge graphs. If the numbers of homeless persons significantly change in a year, details from communities should be obtained. Data relevant to annual updates includes:

- System performance: tied to the Homeless Management Information System (HMIS), which does data cleanup
- National report on trends: obtained from communities with usable data

Other useful data sources and potential issues

Group H contributors identified other data sources essential to the development of the Homelessness OKN and to its utility for end-users after deployment. These data sources, including Veterans Affairs data and data from national research centers studying housing and homelessness, would enable cross-connection between government agencies, such as the VA and HUD, to share data useful in housing homeless veterans and other homeless populations.

VA COVID data

Data about Covid infection among veterans would be of particular use in assessing the effects of infection on housing status. The **Department of Veterans Affairs COVID-19 National Summary**

provides comprehensive information on Covid infection among veterans and its various effects on their lives and livelihood.

Useful data that may be difficult to obtain

The Homelessness OKN could provide greater utility to HUD end-users if data specific to their needs could be obtained. However, initial investigation into this indicates that these data sources may be difficult, if impossible, to obtain. Data that may be difficult to obtain or off-limits to OKN development includes:

- VA "**Homes**" data, which would have to be cleaned before allowing access to it from the Homelessness OKN. Identified data cannot be obtained.
- Data from national research projects that investigate the housing status of veterans. Personally Identifiable Information (PII) data would be used to find researchers and obtain their insights into the causes of veterans' homelessness and possible solutions.
- Data on evictions leading to homelessness, as they happen. Projects like the Eviction Lab at Princeton can help improve access, but has challenges in both gathering the data and handling PII.
- Data on the number of homeless people arrested or imprisoned, or treated in the emergency room, and the aggregate cost of care.

Issues related to collecting hard-to-obtain data

In 2000, soon after HUD launched its Homelessness Management Information System (HMIS), HUD faced backlash from organizations representing homeless populations. Organizations representing the homeless foresaw government agencies possibly misusing HMIS data to track and harass homeless people, especially homeless undocumented immigrants. Concerns included the possibility of domestic abusers accessing HMIS data to locate their domestic partners in shelters to do additional harm.

Finding emergency housing vouchers with an OKN dashboard

Use Case H contributors recommended that the Homeless OKN provide a dashboard for finding emergency housing vouchers available in local communities. The dashboard would be modeled after those used in the city and county of Los Angeles. It would provide for alerts that notify agencies in charge of housing voucher management when leases must be renewed. When lessees cannot afford to renew their leases, agencies can step in to find emergency housing vouchers.

The dashboard would also provide access to Covid infection rates and their effect on the issue of housing vouchers. Property managers may deny leasing to lessees suffering from serious Covid infection. As an excuse, they may say that there is a lack of affordable housing. The lag time between voucher issue and property lookups indicates property managers' reluctance to find suitable housing for this population, even though with housing vouchers, they have the financial means to pay for it.

Case workers for this population could use the dashboard to match people possessing vouchers with available housing. The dashboard would provide information that case workers could use to speed

referrals for homeless people who have emergency housing vouchers but are unable to use them due to a lack of current information about housing availability.

Tying local city planning requirements to data essential to planning for and providing emergency housing would be valuable in OKN development. Dashboard access to city requirements for rental and move-in assistance would be invaluable to case workers and their clients. Data that shows housing availability by city and region could be used by government decision-makers in allocating resources for housing the homeless and for distributing emergency housing vouchers.

On the dashboard, those communities showing the greatest need would receive the largest resource allocation.
Datasets containing information about city planning requirements and the rate of evictions within cities/ defined municipalities could be useful in populating this OKN and providing access to this information through the OKN dashboard.

NOTE: HUD Office of Policy and Development and Research interviewees said that Use Case H developers should match their dashboard requirements to those of city agencies or else ask agencies to change their requirements for housing access and for tracking housing availability. Interviewees recommended that each new jurisdiction contributing data to this OKN be required to use uniform reporting methods, yet to be established.

Identifying essential dashboard requirements

Group H contributors identified possible Veterans Administration contacts whose input into dashboard requirements is essential to the dashboard's efficacy in providing access to useful data. They also identified other dashboard requirements and data sources.

Requirements include opening local data channels to allow OKN access from these channels. This requirement would give users quick access to service, reduce the workload of data collection, and improve service engagement with the OKN.

For example, building a service connection with the Social Security Administration would allow the OKN to ping the SSA database for data on the numbers of persons having disability designations and those who applied for but have not yet received this designation. Data would not identify individuals, but only numbers in each category. Collecting this data would allow it to be aggregated to allow an accurate number of persons with disabilities in municipalities, states, regions, and nationwide.

Contributors noted that The Veterans Administration's **SQUARES** (Status Query and Response Exchange System) provides a model for how the OKN might gather data with short, easy-to-answer questions. Affirmative answers to the question, "Are you a veteran?" are quickly captured and aggregated for further use.

Contributors determined that, at minimum, dashboards would be required to display or provide access to the following information:

- PIT (point-in-time) count : covering vets and others (general **CoC dashboards**)
- **HOMES** data, in the VA : not public.
- HUD **Annual Homeless Assessment Reports**
- HUD **Emergency Housing Voucher** (EHV) Data Dashboard

Policy map

Contributors identified a commercially available policy map as a needed resource for this OKN. This interactive online resource, known as **PolicyMap —Transform Data Into Insights**, maps data at a granular level from communities that users identify and map. The policy map enables users to make informed decisions about homelessness in their communities. The policy map is said to provide up-to-date data essential for insights valuable in decision-making.

Partner OKNs

Contributors identified child welfare agencies and nonprofits, such as Big Brothers/Big Sisters, as possible OKN partners.

Counties

Contributors identified city, state and county Continuum of Care Programs provided through the HUD Exchange as possible data resources for this OKN. Identified COCs include:

- Los Angeles city and county COCs
- Allegheny County, Pennsylvania COC
- **Connecticut Balance of State (CTBOS)**
- **CT balance of state**
- Columbus, Ohio, COC, which provides extensive information on strategic planning

Required resources for a minimum viable product (MVP)

The following resources are needed to develop a Minimum Viable "Product" (knowledge network):

- Existing standard datasets
- **Homelessness Management Information System**, which provides:
  - Point-in-time counts of homeless populations (via Housing and Urban Development)
  - Housing availability (broken down by type and audience)
- Existing schemas/ontologies and research groups, including:
  - The HUD ontology
  - **Observational Health DSI**
  - VA **HOMES** data

NOTE: Existing schemas/ontologies are largely disorganized. Others should be identified and integrated.

- Existing community resources, including:
  - UPenn, Pitt, NYC, LA – all have specialized data and city networks
  - Built for Zero - community-run, for 90 cities
  - Ending Veteran's Homelessness - 80 cities

Group H end-users & stakeholders

Users include homeless individuals and families and city mayors and service providers. The Homelessness Knowledge Network OKN would provide users and stakeholders, identified below, with information specific to their needs. It would provide information difficult to obtain in current siloed systems. It would integrate siloed data to provide users information specific to their needs. It would do so under the circumstances of each user group.

See Current Issues with Data Access for examples of inadequate data access and service provision.
See Mitigation for Negative Social Implications for examples of OKN mitigation of these issues.

Stakeholders

Stakeholders are government agencies charged with providing shelter and services to homeless populations. Community networks and philanthropic organizations are also stakeholders. Examples of stakeholders include:

- U.S. government agencies (HUD, VA, CDC)
- City governments, for example: Aringtonl/Austin, Boston/Somerville, Davis/Woodland, and Seattle/Snohomish
- Community networks, such as: Built 4Zero, VetZero, Mayor's Challenge
- Philanthropic organizations, such as: Bloomberg, MacA, 400 Continuum of Care Programs, and five main HMIS vendors.
- Technology service providers

Current issues with data access

Current systems lack interoperability, making data required for service provision difficult to access. Homeless individuals seeking services may wait longer than necessary to find housing. Human services personnel charged with assessing the homeless individual's eligibility for services spend hours of time confirming the homeless individual's eligibility status and finding funding to pay for the homeless individual's house.

These personnel typically must log on to six different systems to determine eligibility and find funding. Each system (e.g., for mental health organizations and the YMCA) has its own funding source and its own user interface. Navigating these different UIs adds to the burden of finding housing and funding quickly. Current systems do not provide city leadership with information needed for effective housing provision. City mayors seek data that will help them assess:

- The impact that permanent housing may have on preventing future homelessness among previously homeless people.
- The impacts that reliable income and on-site services — provided by government agencies and helping organizations — have on preventing future homelessness.
- The state of mixed-income and racial demographics in housing programs for the homeless. **City First Enterprises** is one resource for this kind of data.

Current systems often do not have the information that city leadership seeks. Group H contributors propose that data be collected from users and stakeholders. Data collectors would ask questions specific to each user and stakeholder group. (See *Questions* below.)

Questions

Group H contributors drafted questions to ask of user and stakeholder groups or to be presented in surveys. Questions for each group are listed below.

Questions for homeless families

These questions elicit answers that will serve a basis for Homelessness OKN development.

- What information has been most useful in finding housing, support, and other social services?
- What information do you need that is hard to get (information is scattered, not mobile friendly, slow to access, etc.)
- How do you usually discover / sign up for housing and services? (by phone, library, etc.)
- What is the hardest thing to manage?

Questions for service providers

Service providers include mayors and city planning offices in addition to service providers for the homeless. Questions for these stakeholders include:

- In what contexts do you work with homeless families, homeless services, and related city services?
- What information sources do you rely on most often? Which are the most useful?
- How many systems do you use? How many of these do you have accounts for?
- What sources do you wish existed or wish were more comprehensive?
- What dashboards do you find most useful for your work?
- What points of misalignment do you encounter among multiple systems?
- Do you have to map misalignment points between systems that use different terminology and standards?

Questions for data providers

Questions for data providers target data required for OKN development. Questions include:

- What information do you gather from or about homeless communities? (You may cite pointers to data sets as an information source.)
- What data sources do you rely on or integrate into your work?

- Which data sets include personally identifiable data?
- Does each data source have a mechanism whereby researchers and social workers can access them?
- Are aggregated versions of data sets made public?

Questions for technology providers

Questions for technology providers elicit answers to be used in technological development of the OKN. Questions include:

- What interoperability data standards are needed but unavailable?
- Which stakeholder organizations lack interoperability standards?
- Which available interoperability standards (e.g., NIEM) could be leveraged for OKN development?
- What groundwork is necessary to prepare for interoperability development?
- How would you use dataset indexing, such as the **Housing Data Portal Index**, to map to common data schema?

Ethics & social implications

The Homelessness OKN has both positive and negative social implications. Positive implications include:

- Software and knowledge graphs that help organize information about community services, including healthcare and **social determinants of health**, housing and well-being, can amplify the efforts of social workers and reduce time-to-service provision.
- Open source software, such as Boston's health and housing data warehouse, provided by **Green River**, can be shared and maintained by many cities. Shared data can be used to improve the quality and integration of social-service tools that must be continually updated to meet changing government standards. Sharing data lowers overhead costs for cities. It also lessens the time necessary to obtain siloed data. With lowered overhead costs and more time available, cities contribute to the shared knowledge of the Homelessness OKN

Negative social implications include:

- Potential pushback for redundancy resulting from multiple systems providing the same data / information.
- Poor organization and findability of data obtained from these multiple systems.
- Challenges from big businesses that benefit from blocks free public access to OKN information.

Mitigations for negative social implications

Group H contributors identified technical and human mitigations for negative social implications.

Technical mitigations

Technical limitations include:

- Mechanisms to avoid sharing private data about homeless individuals.

- Mechanisms to avoid injection of fake or malicious data.
- Support for assertions by different groups validating data.

Human mitigations

Human mitigations include:

- Easy-to-find channels for feedback associated with any interfaces or datasets.
- Channels for coordinating with other linked knowledge networks.

Path to success

The path to success for Use Case H is cross-connection with other OKNs that can provide useful data and are willing to share it. OKNs meeting these criteria would benefit from mutual data exchange. OKNS that would offer the most benefit for data sharing are those that address social services issues. Group H contributors identified other innovation sprint groups that would offer and receive benefit from data sharing through the OKN containing all knowledge graphs. These groups include:

- Group P — Compounding Disasters/Events a.k.a. Wicked Problems Institute. Group P's use case provides requirements for planning and resources when a disaster occurs. Use Case P provides useful information for people who become homeless due to disaster and to the service providers who assist them in their time of need.
- Group E — Connecting Investments in Nutrition Security and Climate. Group E's use case documents requirements for establishing food security and a population's health. Food security and the health of the populace are tightly coupled and directly apply to homelessness. Cross-connection between the knowledge graphs of Group H and Group P would benefit both groups by tracking housing security and the contributing role of homelessness to food insecurity and its effects on public health
- Group B — Integrated Justice Platform and Group M — Decarceration and Reunification from Prison Incarceration to Community Reintegration and Family Reunification. Cross-connection between these groups would stop the negative feedback loop, whereby homeless individuals enter the criminal justice system, become incarcerated, and then return to society, which offers them inadequate support.

  The current feedback loop works as follows: A homeless person is incarcerated in part because he or she resorts to illegal activities to support themselves. The homeless person then becomes a prison inmate. During his/her incarceration, the prison inmate loses any marketable skills and the opportunity to leverage them post-incarceration.

  When the prison inmate returns to society, he or she seldom finds work or housing. The newly decarcerated person is unable to obtain adequate nutrition and suffers declining health as a result. The knowledge graphs of Groups H, P, E, and B should be integrated to provide a full range of services to individuals caught in this negative feedback loop.

Group H also identified future knowledge graphs that should be created and integrated. These future knowledge graphs would address the following issues:

- Health, education, and employment
- Building regulations (per city) – HI.co
- Student + church volunteering networks

Group I: Electronic Consent Services

Group I explored the needs and benefits of electronic consent services that would provide easier access to consent forms required for service. At present, accessing and completing consent forms needed for essential health services can be difficult. Such difficulties result from siloed databases and unreliable completion and routing of forms printed on paper.

An Electronic Consent Services (ECS) OKN would allow cross-connection with other OKNs or databases containing data relevant to each OKN sharing data. It would also allow scanning and electronic distribution of completed paper forms.

Contributors

Group I contributors, listed below, are volunteers representing key stakeholders from academia, government, and industry. Stakeholders have direct experience with or interest in this issue.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Brian Handspicker	Open City Labs
Eric Jahn	Alexandria Consultants
Esther Jackson	Columbia University
Katerine Osatuke	Department of Veterans Affairs
Kenneth Berkowitz	U.S. Department of Veterans Affairs
Luis Amaral	Northwestern University
Paul Wormeli	Wormeli Consulting LLC
Noam Arzt	HLN Consulting
Camille Nebeker	University of California San Diego
Cori Thompson	Drummond Group, LLC

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Jim Pasquale (Kantara)	Kantara Initiative
Kristine McCoy	Stanford Alumni Association
Matt Bishop	Open City Labs

Champions

Group I Champions were Brian Handspicker and Matt Bishop of Open City Labs — Integrated Services for Person Centered Care. Mr. Handspicker wrote the original content for Use Case I and convened and managed Use Case I meetings. Mr. Bishop edited Use Case I and contributed to its development.

ECS OKN serves the public with cross-sector partnerships

As a first step in identifying the ways in which an ECS OKN could serve the public, Use Case I contributors investigated the current state of data and consent access. They found that data sets are held and protected by government agencies and non-government organizations.

Their second step was to explore the potential benefits to the public that electronic consent services could provide. They concluded that by giving the public strong privacy and consent tools, the public would be able to share data. Such tools would also enable data stewards to share data, as permitted by individuals, to help homeless individuals obtain housing and services. Housing homeless people reduces the number of homeless encampments that affect quality of life in cities and neighborhoods.

Their third step was to identify potential users. They determined that the ECS OKN would be utilized by agencies and organizations seeking to share data between information silos. It would serve the public good by enabling improved services to people. For example, it would facilitate program eligibility, which would lessen the burden of providing the same information repeatedly. It would also prevent information drift that occurs when transmitted informally.

Individuals and agencies alike would use the ECS OKN to identify systems that hold their information and privacy preferences, and consent and would allow it to be shared among hospitals, homeless shelters, and other agencies.

Their fourth step was to propose a vehicle for serving the public. They propose building an electronic consent service that would enable the sharing of data between OKNs specializing in health and social service issues. The ECS OKN would be used as a common mechanism to enable trusted, secure, and safe data access that drives cross-sector partnerships. These partnerships would address health and

the social determinants of health. They would advance scientific research in a way that is in accordance with federal and state laws and privacy preferences. **Figure 4** displays the routing of data through this mechanism.

By providing people with a single interface to approve or decline access to their information across agencies, the ECS OKN can build:

- Trust between homeless persons and agency personnel
- Cross-sector partnerships between government agencies and healthcare and social-care providers that store and use personal and/or sensitive information.

Cross-sector partnerships will enable the ontology-driven systems of the OKN to improve availability of data, while providing appropriate protection of sensitive information (which may otherwise be withheld due to law, regulation, policy, procedure, or simply uncertainty by those holding that information).

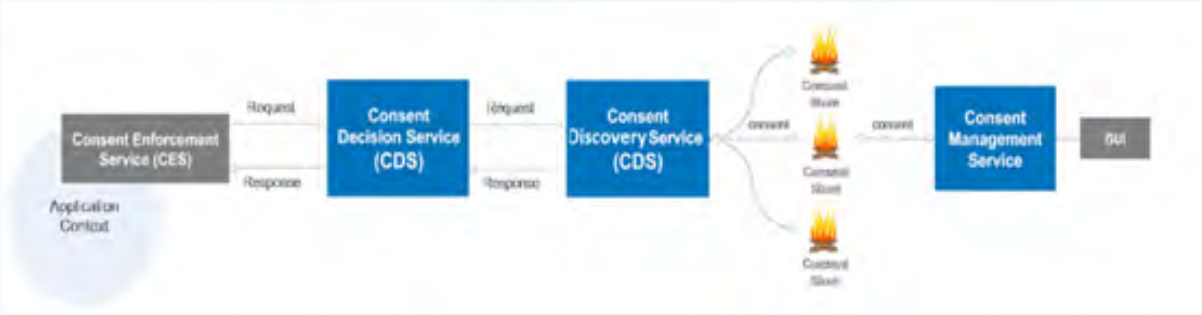


Figure 4: Request and response steps for Electronic Consent Service

Goals & Objectives for the ECS OKN

Using a set of prepared questions, detailed on page 90, Use Case I contributors explored the current state of consent access and the problems that arise from it.

They determined that the goals and objectives for the ECS OKN are to solve the problems arising from inadequate access.

Problems

These problems include:

- Lack of better coordination of care within healthcare agencies, specifically, primary care and behavioral health.
- Lack of hospital and homeless shelter partnerships that would help both hospitals and shelters identify risk factors for homeless people.
- Lack of easy access for homeless people to obtain services across agencies
- Lack of data access for professors and scientists who require data and research participant consent forms to conduct research.
- Legal costs associated with data and consent form acquisition.

- Complex rules pertaining to data access and consent requirements (i.e., when consent is required and when it is not).

Solutions

To solve these problems, the ECS OKN would have to enable:

- Coordination of care between primary care and behavioral healthcare providers
- Hospital and homeless shelter partnerships to identify risk factors for homeless patients and homeless shelter residents,
- Integrated eligibility assessment to provide easy access for homeless people seeking services across agencies

- Easy access to (1) data that professors and scientists need for research and (2) consent forms needed for research participation
- Legal cost funding sources that could be identified through the OKN
- Streamlined rules pertaining to data access and consent requirements
- Streamlined logistics that would enable homelessness research and quicker access to homelessness services. Logistics impeding access to data and to consent recording include:
  - Low adoption rate of open data standards by agencies resulting in:
    - Costly, time-consuming, independent forensic analyses of software systems that use various ways to obtain consent electronically, Forensic analyses determine if a violation has taken place.
    - Inability to identify violators because algorithmic audits become infeasible.
  - Data stewards' misunderstanding of law that prevents data access to people legally entitled to it.

Building a data-access layer in the ECS OKN will meet the following objectives:

- It will support enforcement of state and federal laws and regulations.
- At the same time, it will provide standardized methods for obtaining electronic consent (as required by law)
- Its use will build trust between agencies and the public.
- It will improve access to data for all use cases under development for the Proto-OKN end product.
- It will improve access to consent data even when consent is not required.

Questions to prompt Use Case I development

The following questions guided Use Case I contributors in gathering requirements for development of the ECS OKN.

1) Use/purpose for which we provide consent/permission?
   a) Healthcare example: Use for my care, use for care of people with related conditions, use for quality research. if I die, use it for anything related to healthcare, etc.
2) What are the permissions/consent categories?
   a) By topic? By personal identity features?

   b) Should we enable people to opt-in to particular project(s)
3) At least three levels of granularity for consent to share/access and privacy preferences in healthcare:
   a) Consent/permission for the access to, modification of, deletion, or creation of a document or record
   b) Narrower consent/permissions associated with certain types of sensitive records (mental health, substance abuse, certain diseases, etc. — things could lead to unemployment, discrimination)
   c) Fine-grained privacy marking of individual fields in records (to give atomized consent about what can be shared and to whom)
4) Must take into account the formal constraints such as laws, regulations, and policies of individual organizations, jurisdictions, etc., creating updated and automated procedures for compliance with those constraints
5) To calculate consent we need to know who wants to access the data, the organization they represent, their role, and what they want to do with the data.
6) People should be able to change their mind (about future uses of their data),
   a) We should allow for permissions to expire (and potentially renew) over a set period of time
   b) For research applications of data, expiration of permissions may not be realistic
7) Audit and accounting - people should have the right to know how their consent directives are used and by whom.

How the ECS OKN will be used

Use Case I contributors identified how the consent system/ECS OKN would be used, if built. It also identified probable users.

ECS OKN uses

This consent system would be utilized by agencies and organizations seeking to share data between information silos. It will serve the public good by enabling improved access to services and by reusing data to streamline eligibility and enrollment.

EKS OKN users

This content system would be used by individuals to consent or decline consent and to identify systems that hold their information.

Individuals may choose to share their information when they judge that it will improve the efficiency and/or the quality of the services that they could obtain from an agency or organization. They can choose to withhold information when they deem that the requested information is irrelevant could cause harm if shared.

Individuals could use the consent system to allow their personal data to be used in scientific research that may contribute to the public good.

Challenges and solutions

Group I contributors identified the following challenges to the development and use of the ECS OKN.

- Lack of Consent Ontologies/Data Standards for most non-health domains (NIEM, Human Services, etc.)
  - Identify Use Case Target Services,
  - Select associated Domain models
  - Define Ontologies

Resources

Use Case I contributors identified resources available from health IT standards that could be applied to other domains. These include:

1) **HL7** (Health Level 7), an organization that creates (1) healthcare messaging standards and (2) standards for the representation of clinical documents. It also:
   a) Segments privacy data for the **CDA** (Clinical Document Architecture) and **FHIR** (Fast Interoperability Resources), which are both HL7 standards.
   b) Provides an architecture for modularizing consents and permissions for different areas and use cases
2) **HL7 FHIR Consent Resource**
3) **HL7 Consent Services Functional Specification** (attachment) prototyped by **LEAP-CDS** (LEAP Consent Decision Service) and **Consent2Share**, piloted by **San Diego Health Connect** and **SAMHSA** (Substance Abuse and Mental Health Services Administration) of the U.S. Department of Health and Human Services

They also identified resources to be obtained. These include:

- Ontologies of consent stores (e.g., **FHIR Server consent stores**, **SIF** (Sensitive Instrument Facility) Server consent stores)
- Rules languages for those consent stores (e.g., **XACML**, eXtensible control markup language and **DS4P**, data segmentation for privacy),
- A scalable way to translate laws and regulations into code that enforces consent rules
- Exemplar use cases (e.g., Reunification Use Case M — Decarceration and Reunification from Prison Incarceration to Community Reintegration and Family Reunification)
- APIs for requesting/managing consent (e.g., FHIR Consent, Cardea, Magellan)
- Test systems that are willing to implement the "calls" to the service, retrieve the results, and incorporate them into their workflow.

Other expertise

Use Case I contributors identified areas of expertise that would add to development and result in a robust system.

They identified as essential the following areas of expertise and the names of persons with expertise in specific areas. Names are left blank for those areas of expertise where no human resource was identified.

- Domain-specific Privacy/Consent Lawyers (e.g., with expertise in **HIPAA**, Health Insurance Portability and Accountability Act, and **FERPA**, Family Educational Rights and Privacy Act)
- Domain-specific Ontology experts (e.g., healthcare, human services, justice, EPA), including:
  - Prison Social Work and Administration
  - Criminal Justice System
  - Juvenile Justice (e.g. Decarceration user story of incarcerated teenage single mother)
- Healthcare Coordination - Dr. Kristine McCoy
- Continuum of Care Coordination
- Homeless Management Information Systems — Eric Jahn (a software consultant and vendor representative, not an HUD employee)
- socialcare Case Management — Brian Handspicker
- Child Welfare
- Housing
- Food Social Services (SNAP, WIC, etc.)
- Transportation
- Employment training, career counseling, job placement
- Civil Legal Aid (specifically, legal guidance may be required when data falls under multiple laws and the laws/regulations conflict.)

Stakeholders

Use Case I identified probable stakeholders who would support OKN development and deployment. These include:

- Agency operational staff
- Agency technical staff (internal or contracted)

Persons, especially those identified as "vulnerable" and whose information is held within these systems, may also be considered stakeholders.

Group J: OKN Decision Support for Government

Group J contributors explored the need of government agencies for an OKN that would support decision-making at the federal, state, and local levels of government.

Contributors

Group J contributors are volunteers with knowledge and expertise in the area of data collection and integration. They come from the following organizations:

- Government: 2 (DoT and VA)
- Higher education: 3 (Northwestern, Notre Dame, and UFL)
- Nonprofit: 2 (U.S. Ignite)
- Industry: 3 (Maxar, Infoloom, IBM)

Contributors' names and organizations are listed on the next page.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Michel Biezunski	Infoloom — Integrated Data Curation
Jose Fortes	University of Florida
Tim Goliver	U.S. Department of Veterans Affairs
Nicholas Maynard	US Ignite
Matthew Webber	Notre Dame University

Champion and authors

Group J champion Glenn Ricart led Group J sessions. Group J authors assisted in this. Their names and organizations are listed below.

CHAMPION'S NAME	CHAMPION'S ORGANIZATION
Todd Bacastow	Maxar
Matt Bishop	Open City Labs
Oktie Hassanzadeh	IBM
Murat Omay	U.S. Department of Transportation
Glenn Ricart (Champion)	US Ignite
Amanda Stathopoulos	Northwestern University

Group J: OKN provides strategic decision-making support

Use Case J contributors propose that a Data Integration and Prediction Model be built to support decision-makers at the federal, state, and local levels of government. The proposed OKN would collect data stored in government agency databases and integrate it to support data-based government decisions that have a major impact on the public.

As proposed, the OKN would be a strategic decision-support tool that:

- Provides data-driven insights to decision-makers
- Predicts the range of expected impact of any given decision

This tool would take into account:

- Community capabilities
- Short-term and long-term financial and social costs, and
- Impacts on equity, privacy, transparency, reproducibility, and explainability

It could support decisions in the following areas:

- Land-use
- Transportation options

- Clean air and water regulation
- Emergency preparedness and disaster response

Group J goals & objectives

Group J collaborators identified issues having major impacts on the United States and its citizens. They set as their goal identifying the most effective means for mitigating these impacts and determined that it would be a Data Integration and Prediction Model for a Decision Support OKN. They set as their objectives identifying the best uses of available data residing in silos. Best uses would address the many challenging issues facing the United States. These issues are described below.

Group J issues: Siloed data results in poor decisions

The United States has more than a half-million elected officials and more than 18 million public servants working at all levels of government. These officials and government workers are alone charged with making decisions for the public good. Often, they have little objective data available to make those decisions, leading to arbitrary or decisions driven by interest groups.

With the proliferation of misinformation online, government agencies need a trusted source of information that can be used to shape policies and assess their impact. Government agencies make policies that shape the lives of all Americans.

Good governance and democracy itself depend upon providing governmental decision-makers with linked data that would be available in an OKN. Data residing in siloed databases is sufficiently objective to provide trustworthy information about a number of issues potentially harmful to the American people and environment. The proposed OKN would address these issues, including:

- Geographic (for example, areas most affected by climate change)
- Demographics (important for equity purposes)
- Transportation (affected by limited access in underserved cities and neighborhoods, extreme weather events, and infrastructure degradation)
- Crime (new imagination and economic development may be needed)
- Economic activity (important for equity and planning)
- Air and water quality (important for health impacts)
- Access to health care (known to be an important determinant of personal health)

An OKN that incorporates multidimensional data would drive better policies pertaining to these issues and have profound effects on the health and well-being of the U.S. population. For example, health and human services data identify the social determinants of health. An OKN would use data about these determinants to provide government agencies with relevant and reliable information to shape public health policy. It would help government agencies understand the complex interrelations of policies and their impacts on individuals and families.

The OKN would include sets of issues possibly relevant to decision-making. Decision-makers, particularly in local governments, would have the option of considering data tangential to the problem they are trying to solve. Through data integration, the OKN would be able to identify information relevant to problem solutions that decision-makers would not have otherwise considered.

In an ideal scenario, the OKN could also forecast/predict "What If?" scenarios for various solution options.

Resources

Simple models relevant to governmental decision-making can be used effectively as a resource for understanding the complexity of decision-making as it pertains to decisions affecting public policy and the public's health and wellbeing. Figure 5 displays a decision-making cycle that can guide Data-Driven Decision Support OKN Development.

Impediments to effective data-driven decision support

- Poor data quality could result in poor decisions.
- Conflicting data in the OKN could result in ambiguous predictions of decision impacts.
- When predictions depend upon multiple data variables, it can be hard for humans to search the decision-space effectively. This effect can be mitigated by decision-support systems that allow the decision-maker to see the various predicted outcomes for the local maxima.

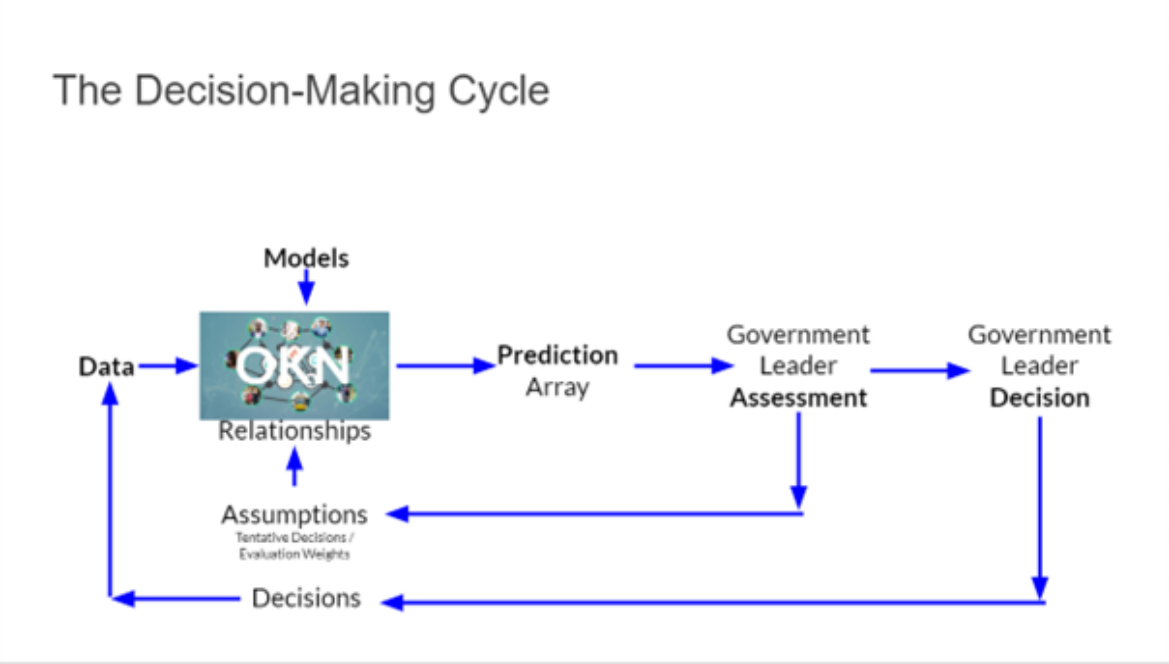


Figure 5: An OKN model for government agency decision making

Benefits of effective data-driven decision support

Excellent data quality and models will yield accurate predictions and grow trust in the decision-support system.

Decision-makers will be better able to explain why they made certain decisions and the outcomes for which they optimized.

Iterative decision-making over time can be used to improve the modeling.

Predicted outcomes can be matched to actual outcomes to improve the models and/or data sources.

The interactive iterative process lends itself to AI techniques for exploring the parameter space (the ranges of possible decisions along multiple dimensions).

Cross-domain decision-making

The DataDriven Decision-Support OKN is intended to support governmental decision-making for cross-domain applications, for example:

- Land use decisions (locations for community anchors, industry, and housing)
- Air quality decisions (emissions testing, zero-emission incentives, measurement, etc.)
- Taxation choices (sales tax, real estate tax, income tax, use tax, and tourism, etc.)
- Choosing where to spend BEAD (Broadband Equity, Access and Deployment) money for connecting the unconnected to the Internet
- Transportation choices (traffic calming, bicycle lanes, and enforcement, etc.)
- Business incentives for increasing employment (tax abatement, and land grants, etc.)
- Expected impact of regulatory decision (e.g., spectrum decision by the FCC)
- Preliminary categorization of simple proposals (e.g., SBIR proposals)

Leveraging existing resources for the Decision Support OKN

Four existing resources could be used effectively to build the Decision Support OKN. These include:

1. Existing GIS (Geographic Information Systems), such as those powered by ESRI for municipalities do exist but are complex, and expensive, and require dedicated staff to populate and operate. It may take months to gather and input the relevant information for a land-use decision or transportation plan. ESRI's ArcGIS software includes a degree of modeling and prediction for limited scenarios but is quite expensive in licensing cost and staff cost.

2. Open source software or online mapping platforms such as QGIS and OpenStreetMap (OSM) may be important building blocks for this OKN project. QGIS is able to provide interactive geographic visualization of model outputs.

3. A wide variety of geospatial (or geospatially referenced) datasets collected and published by federal

agencies, state agencies, municipalities, non-profits, and some companies could be used. But they are often not consistent in content, meaning, or format. These datasets consist of raster (imagery), vector (points, lines, and polygons), and other text-based information sources such as surveys.

4. Standards efforts like the Gravity Project have defined health concerns, diagnoses and procedural and programmatic interventions across a wide variety of domains that touch on programs across multiple agencies.

Finding and creating resources that don't exist

The proposed Decision Support OKN does not have a government sponsor with an interest in OKN decision-making. One would have to be found. Several possibilities were represented on the team (VA,

Best-practice decision-making procedures for the specific pilot area(s) would need to be studied for OKN development.

A panel of decision-makers willing to be the pilot subjects would have to be recruited. Recruiting from a federal agency is recommended.

Existing datasets that support the pilot area would have to use compatible terminology.

A bibliography of relevant academic research and industry practice in the selected pilot area must be created.

Other expertise and resources

Group J contributors identified resources useful to the development of a Decision Support OKN. These include academic papers that could help guide portions of the OKN datasets, prediction models, and interactive visualizations.

They also drafted questions that, when answered, would guide Decision Support OKN Development. These are listed below.

Q1: What are the **most impactful decisions** made by government leaders which lend themselves to being data-driven (by data which is available) and predictable by proven models?

Q2: What **existing examples** are there for data-driven decision-making on which we can build?

Q2: What **existing and emerging datasets** are relevant to government leaders and their decision-making?

Q3: What **proven models** are available for **data-based prediction** in the government space?

Required resources for a minimum viable product (MVP)

Group J contributors have identified possible resources for a Decision Support MVP but have not yet confirmed them. Among required resources are a government sponsor for Decision Support prototyping and datasets and tools, detailed below.

Group J contributors have not yet chosen a pilot area on which to focus prototype development. Their choice depends on whether they decide to merge Use Case J with the use case of Group G — Enabling Community Climate Action and Local-Federal Intelligence Transmission and whether Group G agrees to this merger. If Group J and Group G agree to this merger, they would probably choose a climate-related pilot.

A detailed model for carbon and other greenhouse gasses may not exist. Group J contributors determined that the pilot could proceed by using existing air-quality data. Using this resource would enable pilot/prototype OKN development. The prototype could be further developed into an OKN that could easily transition to carbon-based and other greenhouse gasses.

An example of a prototype that could produce an MVP follows. The prototype would evaluate the impact on air quality at a block-by-block level. The evaluation would target the impacts of various kinds of transportation, housing, and industrial policy decisions.

For this example pilot area, the sponsorship of the EPA is required. EPA sponsorship is required for both the climate action air constituents and the currently more easily measured air pollution constituents.

Datasets

Datasets required for a Decision Support OKN MVP are as follows:

1. Detailed air quality microclimate data from organizations like **Purple Air | Real-time Air Quality Data** and **Mendeley Data Tellus dataset** and from
2. wind/weather readings from the Weather Underground. (This seems feasible already.)
3. Air quality "impact data" from various kinds of transportation (possibly from EPA and/or DOT), housing (possibly from HHS), and industry (possibly from EPA or Department of Commerce or NOAA).
4. 3D data for terrain and the built environment The data would apply to the pilot area and would be suitable for the microclimate model.
5. Details of permitted air pollution sources.

Tools

Datasets required for a Decision Support OKN MVP are as follows:

1. QGIS or ArcGIS
2. A microclimate airflow model that can take in the 3D terrain data and detailed prevailing wind information from Weather Underground The model would be able to infer sources of air pollution

from the air-quality microclimate data and permitted air pollution sources.

3. A decision-support tool friendly to QGIS or ArcGIS

Resource unknowns

What resources could be provided by our federal partners?
Climate change and/or future changes to air quality are unknown. Partnerships with the EPA, the DOE, and/or NSF atmospheric sciences would probably be necessary to obtain and analyze climate change and air-quality data. Whether these government agencies would agree to partnership is unknown.

Evolving resource requirements

Resource requirements must be updated and met as new climate-change and air-quality data become available. Existing resources at the federal, state, and local partner levels would be leveraged to obtain updated data. Publication Resources

Publications that serve as information resources for this OKN are listed below.

Data Innovation Fund project, "Urban Planning Tools as Agents of Change: Collaborative Spatial Data for Sustainable Urban Development in Indonesia"
This publication covers augment planning and service delivery through systems and tools that facilitate a consultative, inclusive, and efficient process. From **The World Bank — Urban planning tools: Collaborative spatial data for sustainable urban development in Indonesia**.

End-users & stakeholders

Data and analytics literacy programs are needed to ensure that the workforce at all levels have the necessary skills to create, comprehend, and/or work with data.

A Decision Support for Government Leaders OKN would serve the public good by enabling better policy decisions that directly affect the public. Decision-makers at the federal, state, and local levels of government would have ready access to OKN data and tools leveraging knowledge graphs. Decision-makers could access information specific to the problems that they must address.

The Decision Support for Government Leaders would expand the ability of stakeholders to find targeted solutions for problems in their communities or in the nation as a whole. This OKN would also allow stakeholders to share strategies for mitigating shared challenges. It would build on existing resources for government decision-making and thus add value to the services that governments provide. Every decision made will, over time, add to the data available to future decisions.

Increasing stakeholders' ability to solve problems

Group J contributors identified three types of stakeholders who would benefit from access to a Decision-Support OKN. Each type would use the OKN in different ways, which are described below.

- Stakeholder type 1 comprises elected officials and civic administrators who use the OKN to accelerate and improve the ability to scale up successful innovation. Data-driven approaches and transparency enabled by the OKN facilitate testing and validation of civic strategies. It also excludes intuitive but wrong inferences. This OKN provides an improved capability to capture and capitalize on spill-over effects of a variety of connected markets, such as employment/labor.

- Stakeholder type 2 represents collaborative decision–makers for MPO/DOTs/Cities. Stakeholder type 2 uses the OKN to share strategies across city/state/administrative boundaries and to address shared challenges (water, transportation, emergency).

- Stakeholder type 3 represents citizens. The OKN can enable citizens to be actively involved in solving problems in their communities. The OKN provides them and the other types of stakeholders with data obtained through crowdsourcing and feedback from community involvement.

Citizens may have more time available to more thoroughly explore the solution spaces and suggest great solutions to their elected officials and community administrators. They can use this data to suggest actions which the OKN model shows may improve the efficiency of services, equity, and buy-in.

Stakeholders

Stakeholders should include those governmental agencies at the local, state, and federal level with responsibility for the target area (perhaps air quality). The project should seek resources and knowledge from MPO's, DOT's, and cities that have investigated or implemented digital and data-driven decision literacy programs within their organizations. The California Department of Transportation (Caltrans) is one example of a state government agency that developed a data-literacy program.

An example at the federal level is the DOD. The DOD established the following requirement for its Intelligence professionals:

"Intelligence professionals [should be] enabled with baseline digital literacy and access to the digital infrastructure and software required for ubiquitous AI integration in each stage of the intelligence cycle. The human talent deficit is the government's most conspicuous AI deficit and the single greatest inhibitor to buying, building, and fielding AI-enabled technologies for national security purposes."

From the Final Report National Security Commission on Artificial Intelligence (NSCAI, 2021).

Human-centered design

Group J contributors will investigate the issues surrounding policy decision-making. Phases include planning, management/operations, and sustainability of policies/infrastructure. They will employ HCD

user research techniques, such as interviews with stakeholders and end-users to obtain their design requirements. Interview questions include:

- What role does data/digital literacy play in each phase?
- Where is there a gap or need?
- What is the highest priority to address?

Ethics & social implications

Group J considered use-case template questions and arrived at the following conclusions.

Limitations on data sources will vary depending on the specific government decision-maker being supported. For the air-quality example, EPA data is fairly uniformly specified. It has few or no issues, such as the security of personally identifiable information.

Some potential biases may exist. Biases may be introduced into the OKN from unintentionally biased data. For example, such data may be collected from communities that place their EPA sensor in a location expected to have the least pollution. However, the multiple-inexpensive sensor approach used to explore variations within a community will mitigate this bias.

Again, gaps or omissions in data depend upon the government decision-making task attempted. For the air-quality example, any gaps or omissions can be minimized by the use of multiple inexpensive sensors. Use would extend throughout the air-quality community. U.S. Ignite has deployed multiple sensors in four of its air-quality communities. PurpleAir also has deployed an extensive inexpensive sensor network.

Determining who has rights over data and whether permissions are needed depends upon the specific government decision-making task attempted. For the air-quality example, no permissions are needed.

Potential harm to communities depends upon the specific government decision-making task attempted. For the air-quality task, data is more likely to bring existing injustices and equity issues to light. Individuals living in higher pollution areas may receive greater attention. Group J contributors propose that members of all communities be given access to tools and data so that they could create their own "What If?" scenarios.

Engagement by those who may be impacted will be comparatively easy when contrasted with other potential use cases. Public trust is likely to be improved. Air-quality measurements over time will show how well policy and action have evolved to address air-quality concerns.

Path to success

Use Case J is related to all other use cases where the end goals are decision-making by government officials and the insight gained. Citizens can use the Decision Support OKN to model the impact of

potential government decisions.

While Group J contributors have chosen government decision-makers as the likeliest data providers based on the availability of government data and the lack of other available tools, contributors have also made a strong case for the decision-making tools boosting the data-driven decisions of private industry. This is particularly true for private industry contributors to issues of national importance where measurements are already being taken. Some examples include private industry response to aridification, wildfire abatements, and adaptations for climate change.

Sustainability and scalability

Efficient governance has an upfront cost but a lasting benefit. Data collection and models require funding for collection and development. Establishing relationships between government agencies requires time and funding. Models will help decision-makers anticipate the likely impacts of their decisions on the public and provide a way to explain those impacts. Good decisions will be strengthened. Poor decisions will be more easily second-guessed.

The Decision Support OKN would automatically be sustained if benefits from its use continue over time. If the first few knowledge graphs contributing to this OKN yield substantial benefits, the OKN would be scalable with available data.

Encouraging journals to request decision-support models could support sustainability. Journals could request decision-support models that support research. Journals could request such models in their supplementary material submissions.

The NSF could request data for research and make it openly available to the research community. The NSF could also use requested research data to create decision-support tools that allow decision-makers to ask “What if?” questions.

Other concepts & issues

Some kind of “insurance” backing the OKN data would be helpful. That “insurance” could come in the form of one or more risk-takers who back the data (as in title insurance for property deeds). Alternatively, multiple sources of data would be another way to provide confidence in the data.

Group K: Collaborative Knowledge Network

Group K developed a use case that captures the requirements for a Collaborative Knowledge Network (CKN). The proposed CKN would provide a persistent and accessible environment to OKN users and stakeholders. Key requirements for the CKN are that it protect users' privacy and allow users to share data freely.

Contributors

Group K contributors are volunteers from academia, government, and industry. Each has expertise in subjects relevant to development of Use Case K. Contributor names are listed below.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
He Bai	Oklahoma State University
Carl Busart	Army Research Lab
Lauren Close	Columbia University
Nick Del Rio	U.S. Air Force
Jemin George	Army Research Lab
Florence Hudson	Columbia University
Quyen Nguyen	U.S. Census Bureau
Emily Rothenberg	Columbia University
Greg Seaton	SierraLogic
Ilya Zaslavsky	U.C. San Diego

Champion

Greg Seaton of SierraLogic was champion for Group K. As champion, he organized and led Group K meetings.

CKN protects privacy, allows data sharing

The proposed CKN will serve the public good with an accessible and persistent platform for collaboration. It will have built-in privacy protection and security for data sharing. Privacy and security features will be based on declarative data policies and governance. These features will enable users to share, discover, and leverage knowledge. This safe, collaborative environment will enable users to gain understanding of the world and its various communities and domains. With their newly acquired understanding, users will be able to collaborate to develop strategies for effecting change.

Beneficial societal impacts of the CKN

Group K contributors foresee significant societal impacts resulting from CKN use.
These include:

- Mitigating systematic patterns of abuse and neglect identified by the CKN.
- Per FAIR principles, making knowledge easily accessible and tractable
- Expanding knowledge exponentially through building and sharing enabled by the CKN and modeled after the growth of global communications enabled by the Internet
- Democratizing knowledge building, thus preventing an authoritarian monopoly on knowledge
- Enabling collaborators to stand-up their own CKN node at their university, company, government agency, and/or non-government organization.
- Leveraging community CKN nodes for zero-infrastructure participation; for example, community CKN nodes could host collaborator knowledge graphs (private, protected, public) and provide access to the CKN with minimal/zero IT resources. (See **figure 6** for a depiction of collaborative node structure.)
- Leveraging and supporting current and future OKN efforts (e.g., OKN Infrastructure, KnowWhereGraph, SCALES).
- Providing the infrastructure to publish, access, and consume collaborator graphs with managed vocabularies and ontologies.
- Enabling collaborators (institution, organization, team, or individual) to control what is published and who has access, etc.
- Providing limited but critical governance and oversight of:
  - Network communities and domains, and
  - Overall network operations and administration

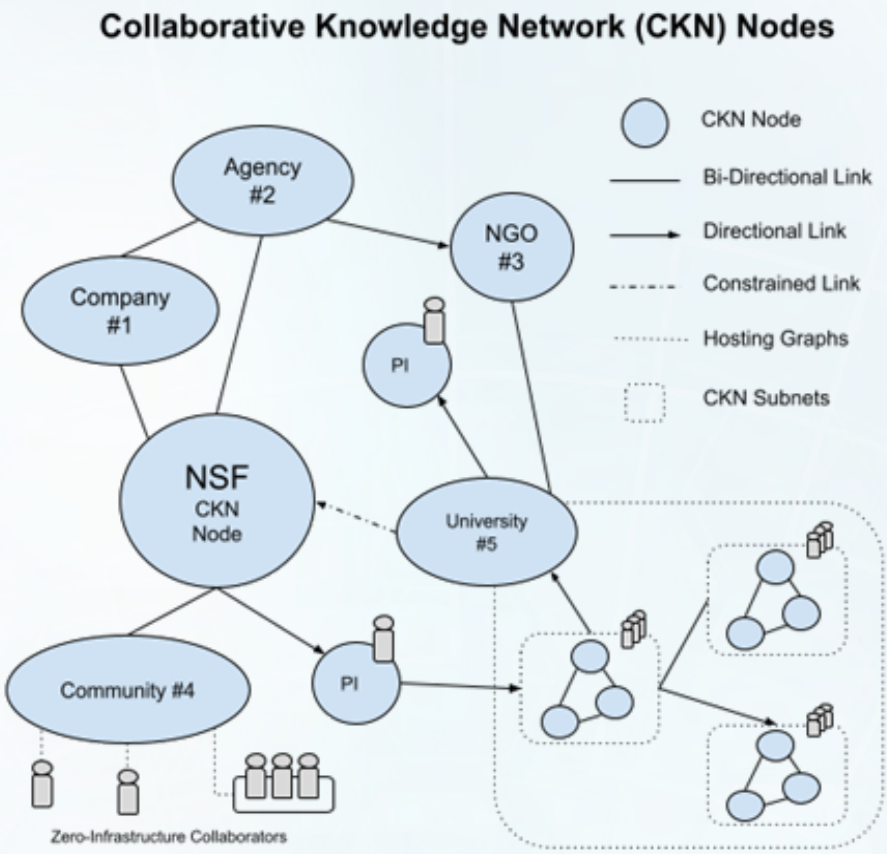


Figure 6: CKN Node Structure

Goals & objectives

Group K collaborators determined the goals and objectives for Use Case K development. They agreed that Use Case K must:

- Provide a policy platform to manage data policy.
- Support horizontal integration of disparate datasets, resources, and knowledge graphs to coordinate policy development and planning.
- Support distributed inference and reasoning across knowledge graphs for implicit pattern recognition.
- Provide semantic knowledge graph hosting for zero-infrastructure support.
- Enforce data privacy, ethics, and security policy across knowledge graphs/networks.
- Provide data consistent with FAIR. principles (Findable, Accessible, Isolated, Reusable).

Resources

Group K contributors identified the following resources:

- Sample datasets in domains that require controlled access, including:
- Clinical trials, financial, intelligence, law enforcement, sociological research
- Sample datasets with simple data access and control policies

- Sample datasets with moderate data access and control policies
- Sample datasets with complex data access and control policies
- Existing models for CKN model/ops areas to possibly evolve/extend

Other expertise

Group K contributors identified domain SMEs, who could provide a larger use-case spectrum.

End-users & stakeholders

Group K contributors identified CKN users and stakeholders. Each group is listed separately below.

Users

Identified users include:

- Principal Investigators
- Researchers
- Domain Community Leaders
- Domain Communities
- IT Departments
- General Public

Stakeholders

Group K identified stakeholders needed for CKN development but not yet or recruited. They determined that stakeholders must be willing and able to build out the CKN. They also must be able to:

- Describe and enforce data policies in a shared knowledge environment
- Share/publish their community / domain / organization datasets/graphs, and
- Leverage knowledge from external authoritative sources with funding, authority, and resources to contribute to CKN development.

Human-centered design

User needs and preferences contribute to the human-centered design to be used in CKN development. Group K contributors identified the following user and stakeholder requirements for a human-centered design of the CKN. The CKN must:

- Enforce data policies across datasets / graphs in an accessible and persisted environment based on standards.
- Query federated data sets / graphs in an accessible and persisted environment based on standards.
- Publish datasets / graphs to an accessible and persisted environment based on standards.

Stakeholders

Group K also identified stakeholder requirements for human-centered design. For stakeholders, the CKN must:

- Define data policies declaratively to enforce permissions and access controls in an accessible and persisted environment based on standards.
- Publish organizational and community datasets / graphs in an accessible and persistent environment based on standards.
- Leverage knowledge from distributed organizations and communities in an accessible and persisted environment based on standards.

Ethics & social implications

Group K identified both the beneficial and detrimental implications of CKN use.

Beneficial implications

The CKN may have significant beneficial impacts, listed below.

- The CKN could identify societal and systemic patterns of abuse/neglect that should be addressed.
- The CKN could make knowledge easily accessible and tractable (FAIR principles).
- The CKN could bring about an exponential growth in knowledge building and sharing, as the Internet has brought growth in global communications.
- The CKN may promote democratization of knowledge building, thus preventing any monopolized authority of knowledge.

Detrimental implications

The CKN may have detrimental impacts, listed below.

- The CKN may have a detrimental social impact if the collaborative knowledge is used to suppress/punish racial groups, gender identification, sexual preferences, and cultural, political, and/or religious groups.
- CKN may have a detrimental social impact if malicious data/graphs are injected into the OKN.
- CKN may have a detrimental social impact if sensitive and/or private information is intentionally or unintentionally published.

Mitigations

Group K identified possible mitigations for CKN misuse. They include mitigations for both technological and human errors.

Technology Mitigations

Group K collaborators established the following requirements for technology mitigations.

- The CKN requires mechanism(s) to mitigate the abuse of sensitive and private data.
- The CKN requires mechanism(s) to mitigate the nefarious use of OKN data once identified.
- The CKN requires mechanism(s) to mitigate the intentional injection of malicious data/graphs.
- The CKN requires mechanism(s) to mitigate monopolized authority over data to support democratization of knowledge via decentralization.
- The CKN requires mechanism(s) to allow 'certification' of datasets/graphs using verifiable credentials (claims), or some other cryptographic non-repudiation approach, to ensure the datasets/graphs have been reviewed and originate from reputable sources.

Human mitigations

Group K collaborators established the following requirements for mitigation of human error.

- Community members must identify, document, and report abuses in the OKN/CKN.
- Communities must review mechanisms to ensure that private and/or sensitive data is not published publicly intentionally or unintentionally and is not accessible by unauthorized parties.
- OKN, CKN, and community leaders and management must support efforts to address abuses and communicate with other leaders, managers, and members to not only identify individual isolated cases, but to also identify and mitigate patterns of abuses across the OKN/CKN.

Path to success

Group K collaborators identified requirements that the CKN must meet to be successful. CKN requirements are listed by category below.

General requirements

Group K collaborators established the following general requirements for the CKN.

- The CKN will handle graph and linked data using W3C standard RDF, RDF/S, OWL, and SPARQL.
- The CKN will handle graph and linked data using JSON-LD.
- The CKN will support 'zero-infrastructure' network node installations and configuration with minimal/no IT infrastructure requirements and minimal IT expertise.
- The CKN nodes will provide SPARQL 1.1+ (or a pragmatic subset) endpoints.
- The CKN will provide mechanism(s) to extend functionality via modular services.
- The CKN will provide hub node centralization for published graphs and datasets.
- The CKN will provide decentralization of nodes into a mesh of knowledge nodes.(See **figure 7**).
- A graph owner can request to join the CKN.
- A graph owner can update the knowledge nodes of his or her graph.
- A CKN user can propose updates to knowledge nodes.
- The CKN will provide a voting process to approve the proposed updates.

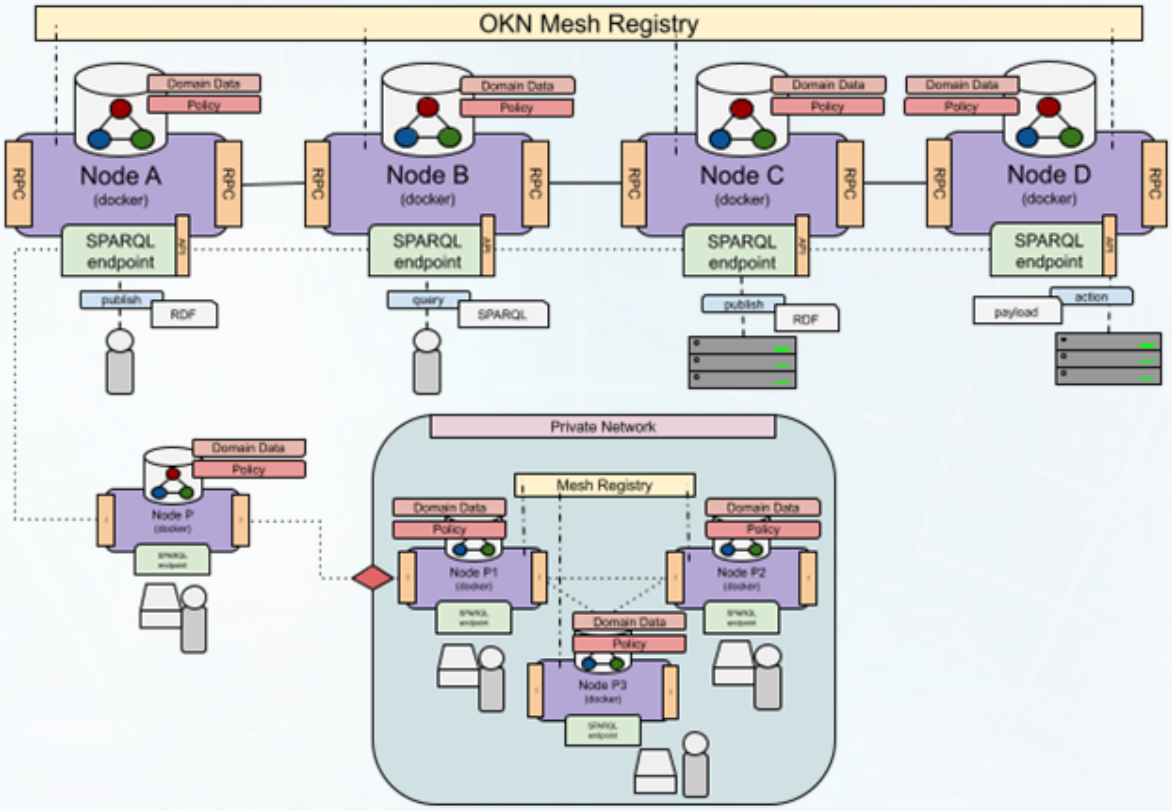


Figure 7: Decentralized nodes in an OKN mesh registry

Publishing requirements

Group K collaborators established the following publishing requirements for the CKN.

- The CKN will provide the ability to publish/pin decentralized graphs and datasets via decentralized identifiers.
- The CKN will support publishing in RDF.
- The CKN will support publishing in JSON-LD.
- The CKN will support secure data file escrow services for exchanges of data files across the network.
- The CKN will provide a voluntary knowledge registry for browsing and automated discovery of domain data across the CKN and external sources.

Querying requirements

Group K collaborators established the following querying requirements for the CKN.

- The CKN will support querying across decentralized CKN nodes via SPARQL. (See **figure 8**.)
- The CKN will support inference across decentralized CKN nodes.
- The CKN queryable nodes will expose SPARQL 1.1+ (or a pragmatic functional subset) endpoints.
- The CKN will enforce data policies and permissions on all queries.

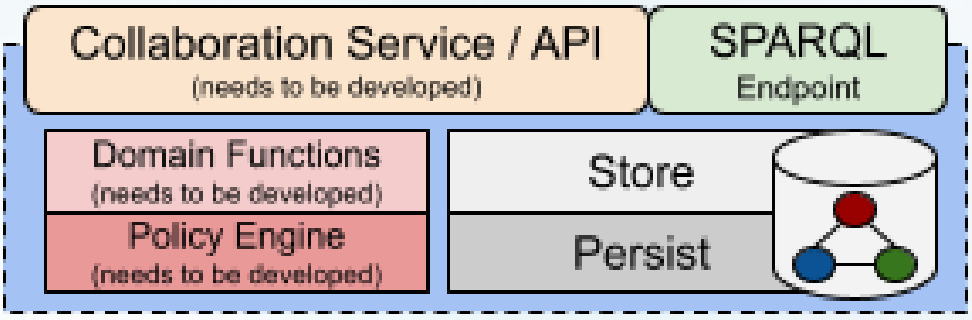


Figure 8: Internal node components of CKN

Policies and permissions requirements

Group K collaborators established the following policies and permissions requirements for the CKN.

- The CKN will provide mechanisms to declaratively define data policy on nodes, networks, subnetworks, and data elements.
- The CKN will provide mechanisms to control access and permissions to individual nodes.
- The CKN will provide mechanisms to control access and permissions to networks and subnetworks of nodes.
- The CKN will provide mechanisms to control access and permissions to individual graphs and datasets on individual nodes.
- The CKN will provide mechanisms to control access and permissions to individual data elements.

Security requirements

Group K collaborators established the following security requirements for the CKN.

- The CKN will encrypt all in-flight communications via SSL.
- The CKN will provide a mechanism to encrypt graphs and datasets at-rest (stored).
- The CKN will provide mechanisms for decentralized verifiable identities and permissions.
- The CKN will provide mechanisms for centralized identity management.

Administration requirements

Group K collaborators established the following administration requirements for the CKN.

- The CKN hub and standard nodes will provide mechanisms for viewing node state.
- The CKN hub and standard nodes will provide mechanisms for allowing admins, managers, and other permissioned users to manage node access and permissions.
- The CKN hub and standard nodes will provide mechanisms for allowing admins, managers, and other permissioned users to configure and manage node operations.

Governance requirements

Group K contributors established the following governance requirements for the CKN.

- The CKN will provide a standard technical framework for governance at the node, domain, community, organization, and network levels.
- The CKN will provide enforcement of governance across nodes, domains, communities, organizations, and networks of policies defined by designated oversight roles.
- The CKN will provide governance audit and reporting to support transparency and trust of CKN operations.
- The CKN will provide mechanisms for governance at the data, knowledge, and operations levels.

Connectivity

The Collaborative Knowledge Network initial efforts provide a blueprint to develop the backplane or plumbing to allow for the collaboration of OKN working groups and communities.

Sustainability and scalability

Group K Collaborators envisioned CKN sustainability and scalability for the short term (1-3 years) and long term (4 - 15 years).

Short-term / develop

Short-term requirements for sustainability and scalability are:

- Secure funding to build/develop the first CKN iteration
- Generate a short-, mid-, and long-term operations plan for the CKN

Continuing Sustainability

Requirements for continuing sustainability are:

- General long-term continuing oversight of CKN
- Maintenance, bug fix, and new features
- Domain/subdomain data governance oversight
- Ethics oversight/review/reporting
- Privacy/Consent oversight/review/reporting
- Cybersecurity oversight
- Outreach and education

Other concepts & issues

Group K contributors identified the following additional concepts crucial to the success of the CKN.

- **Data Policy:** Leverage existing and new ontologies and work to provide the data policy OKN ontology that will support a model for declarative data policy based on identifiers and group associations.

- **Data Permissions:** Leverage existing and new ontologies and work to provide granular access and rights to individual data elements.
- **Operations and Administration Policy:** Leverage existing and new ontologies and work to provide an operational and administration policy OKN ontology that will support a model for declarative operational and administration policy based on identifiers and group associations.
- **Operations and Administration Permissions:** Leverage existing and new ontologies and work to provide granular access and rights to individuals, automated actors, services, and organizations to operate and administer access and other permissions.
- **Knowledge Network Operations:** Develop ontologies to support operational and transactional data for day-to-day operations of a knowledge network mesh to provide the necessary data/audit to run the CKN backplane (e.g., how to represent nodes in the network, node capabilities, node and graph registries, node subnets, etc.).
- **Long-Term Operational Plan:** (what oversight is needed, ethics reporting/resolution, mechanism for bug fixes, etc.)
- **Functional Requirements:** (from a business perspective, etc.)
- **Non-Functional Requirements:** (what is expected from a UX perspective, etc.)
- **Technical Requirements:** (how restrictions for technologies, etc.)
- **Preliminary Technical Specifications:** (how to actually build CKN; blueprint)

Group L: OKN for Energy Management
Achieving 100 Percent Renewable and Resilience

Group L focuses on the development of a decision support tool that would allow government agencies, utilities, and local communities to assess the risk to the power-grid from extreme events. To reduce the energy burden, the decision-support tool would also determine community energy resilience and identify potential uses for renewables.

Contributors

Group L contributors include representatives from academia, government, and the private sector. Contributors to Use Case L and their organizations are listed below

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Bandana Kar	Oak Ridge National Laboratory
Ellie Young	Common Action
Lilit Yeghiazarian	University of Cincinnati
Ryan McGranaghan	NASA
Douglas Rao	NOAA
Matt Gee	BrightHive
Johannes Friedrich	World Resources Institute
Jianxi Gao	Rensselaer Polytechnic Institute
Ted Habermann	Metadata Game Changer
Jennifer Melot	Georgetown University

Champion

Bandana Kar from Oak Ridge National Laboratory was the champion for Group L. As champion, Dr. Kar organized use-case group meetings and briefed contributors on group updates.

Enabling reliable, affordable access to energy serves public good

The OKN for Energy Management — Achieve 100% Renewable and Resilience serves the public good by enabling underserved communities to obtain energy when their access is limited due to rising prices, service disruptions, and extreme weather events.

Group L contributors have clearly laid out the problem of energy insufficiency and the problem solutions enabled by Energy Management OKN. They also detail the OKN's contribution to the public good. The problem, contribution to public good, and the solution are described below.

Problem: In the U.S., reliable and affordable access to energy is limited by rising energy prices and aging infrastructures. Significant power outage and service disruptions due to extreme weather events contribute to rising costs. Extreme weather is exacerbated by climate change. Vulnerable communities often prioritize immediate needs (employment, health care, education, public safety) over investing in renewables to address energy access and equity concerns.

As part of the Strategic and Interagency Initiatives (SI2) within the Department of Energy's (DOE's) Office of Energy Efficiency and Renewable Energy, agencies committed up to $355 million to develop data and tools (Low-income Energy Affordability Data (LEAD), Clean Energy for Low-Income Communities Accelerator (CELICA) to assist 155,000 underserved households with energy efficiency. Yet, vulnerable communities are experiencing a higher energy burden and are not resilient.

Public Good: The purpose of the OKN for Energy Management is to enable reliable and affordable access to energy. The targeted users for this OKN are at-risk communities that need assistance in achieving climate resilience. Because one solution will not be applicable for every community or event, stakeholders (FEMA, DOE, utilities, local communities) would benefit from having access to the following resources:

- Datasets pertaining to:
  - Energy infrastructures
  - Energy technologies
  - Energy prices for different technologies
  - Outage data and socioeconomic conditions

Because such data are not readily available and often require data harmonization and fusion to provide situational awareness information, Group L contributors propose building an OKN based decision-support system.

Solution: Our solution is to develop a decision support tool that would allow DOE, utilities and local

communities/stakeholders to assess risk to the grid from extreme events, determine community energy resilience, specifically, energy justice and burden, and identify potential places for renewables to reduce energy burden while achieving clean energy. The OKN will have three parts: (i) connect climate data with energy infrastructure data to assess and improve grid resilience; (ii) connect socioeconomic data with energy grid to identify and improve community energy resilience; and (iii) connect actors (local communities, utilities and stakeholders) to identify strategies to address energy resilience and management.

Goals and objectives

Group L contributors established as goals and objectives meeting the energy needs of underserved communities and supporting them in achieving energy resilience. Group L contributors identified potential users of the Energy Management OKN. They developed Use Case L to document the specific requirements of each user group. Potential users and applicable use cases are described below.

Potential users and use cases

- The DOE can use the system to identify (i) risk to current the grid under different climate conditions and (ii) potential solutions to using clean energy technologies.
- Utilities could use the OKN to assess the risk to their infrastructures, potential impacts on communities, and determine supply and demand relationship under normal vs extreme conditions. Smaller utilities with limited funds and rural user base can use the OKN to identify their investment strategy.
- Local communities can use the OKN to assess their energy resilience (energy justice and energy burden) in the context of climate extremes.
- The OKN will allow integration of disparate datasets and bi-directional communication between stakeholders to address energy resilience.

Resources

Group L identified existing resources for development of the Energy Management OKN. The group also identified resources that need to be created. Both resources are described below under separate headings.

Existing resources

Existing resources are robust and easily accessed. They provide a wide range of data and datasets essential to the development of the Energy Management OKN. Existing resources include:

Foundation layer

The foundation layer is the platform for further OKN development. Existing foundation layer resources include:

- KnowWhereGraph — to be used as framework to locate spatial locations of impacts and features

- Energy System Resilience OKN to identify energy and hazard relationships

Climate data

Climate data serve to predict severe weather that could affect energy access. Climate data also provides information useful in preparing for disasters such as floods that leave people without energy. Climate data useful in OKN development include:

- NOAA Climate at a Glance:
- NOAA Billion-Dollar Weather and Climate Disaster outcomes and their risk potential
- Climate Explorer
- Urban Flooding Open Knowledge Network
- NOAA State Climate Summaries
- EarthCube

Socioeconomic data

The Energy Management OKN would use socioeconomic data to identify communities with the least energy resilience. Sources for socioeconomic data include:

- World Pop
- U.S. Census
- Shared Socioeconomic Pathways

Energy data

Energy data is a key component of Energy Management OKN development. It includes:

- Global Power Plants Database
- Power Outage Data (EAGLE-I)
- Infrastructure Data (HIFLD)

Plans

Plans provide guidance to Energy Management OKN development. Plans include:

- DOE Resilience Plan
- DOE Net Zero Initiative
- World Resources Institute
- DOE — Energy Justice Initiative

Community

This resource is Information obtained from stakeholders and utilities through engagement and news media.

Ontologies

Two ontologies useful to OKN development quantify earth and space science threats to the power grid. They also aid understanding of the existing energy infrastructure. The ontologies include

- The Convergence Hub for the Exploration of Space Science
- Towards a Modular Ontology for Space Weather Research

Necessary to Develop

Group L contributors determined that new domain ontologies must be developed to provide information unavailable in existing ontologies. Ontologies to be developed must cover the following domains:

- Energy sectors
- Interactions between energy infrastructure and extreme events
- Impacts of energy infrastructure failure on communities
- Ability of potential users and stakeholders to address energy justice and burden.

Other expertise

The OKN requires engagement with DOE, utilities, and communities. It requires user experience professionals to assess the effectiveness of the OKN framework with different user and stakeholder groups identified in the Use Case L narrative. To develop an effective OKN, a range of professionals with relevant expertise must be recruited.

The Energy Management OKN developers will need to connect with persons with expertise in grid resilience and renewable energy and researchers with expertise in this area.. The OKN will also need:

- An earth and space scientist to study potential the impacts of climate extremes
- Urban planners and geospatial scientists to assist with planning where renewables should be located
- Social scientists and data scientists
- Operation researchers
- Energy economists and policy experts, and
- Computer scientists to assist with the development/deployment of:
    - Needed ontologies that do not yet exist
    - The Energy Management OKN

Resources for developing a minimum viable product

To develop a minimum viable product (initial knowledge network), contributors will use the open-source datasets, tools and existing infrastructures available (discussed in **Resources** for the Energy Management OKN).

Other resources needed for MVP are:

1. Community Targets:

a. DOE Resilience Plan, DOE Net Zero and Decarbonization Initiatives,
b. Discussion with communities (specifically, rural and remote communities) to develop OKN components.

2. Expert Communities: Data stewards, policy makers and topical experts such as World Resources Institute, Electric Power Research Institute, North American Electric Reliability Corporation, etc.
3. Social infrastructures to facilitate interaction between stakeholders (communities, utilities, planning entities, policy makers) for identification of user needs, data and network components.

In its original version, the focus would be to develop a basic technical infrastructure as presented in **figure 9**. Details for development at each stage follow.

Figure 9: Stages of basic technical infrastructure development

ETL pipeline/Ingestion would include: Access to external data sources, Mapping mechanism for legacy data (relational, semistructured, unstructured, etc.) and Data cleansing facilities

Storage: ostensibly a graph database, triple store, etc., for ontologies and knowledge graphs of the selected use cases and a **Data catalog** and business glossary to assist with data identification and knowledge creation based on specific use cases.

User interfaces: to assist with Curation/administration interface and Exploration interface for end-users

APIs: to be developed based on open standards (W3C). APIs will allow others to build applications on top of the OKN by providing:

- Knowledge graph access
- Ontology management
- Provenance/lineage/etc. administration.

Resources unknown

Group L contributors were unable to identify all the datasets needed both from the public and private domain, as they are dependent on use cases. They were also unable to identify all the potential users and stakeholders for this OKN and its computing needs.

Once the graph is established as part of the OKN, it will help identify:

- Needed data sets
- Ontologies that need to be developed or leveraged
- Computing needs for OKN deployment
- Governance and security needed to ensure:
  - Adherence to ethical standards
  - Security mechanisms are in place

End-users & stakeholders

The proposed Energy Management OKN will have several end-users and stakeholders, each with specific needs. The OKN will serve the public good through the work that end-users and stakeholders perform. Using the OKN, government agencies, private entities, policy makers and others will bring about change leading to renewable sources of energy and energy resilience. Government agencies and other end-users and stakeholders are listed below.

End-users

Energy Management OKN end-users range from federal agencies to private entities and NGOs. End-users include:

1. **DOE:** The tool could be used by the DOE in conjunction with its NAERM and EAGLE-I programs as a decision-making tool to identify (i) risk to the current grid under different climate conditions, (ii) potential clean energy technologies and where they can be used to improve the energy resilience of communities and the grid; and (iii) potential policies to address supply-chain needs for clean energy technology development.

2. **Private Entities:** With a focus on renewables, private entities are spending on the generation of solar panels, wind turbines, microgrids and Distributed Energy Storage systems. The OKN could be used to identify potential markets/locations for renewables as well as supply chain management needs.

3. **Policy Makers:** Policy makers (e.g., local planning agencies) can use this information to identify energy resilience and security policies and programs to increase community adaptation. They can use the tool to identify planning activities that must be undertaken to address net zero, installation of renewables, and expansion of infrastructure under the Infrastructure Investment and Jobs Act.

4. **Climate Action Groups and NGOs:** With DOE's focus on achieving net zero globally, the basic infrastructure of the tool could be repurposed for other countries. Climate Action Groups and NGOs could also use this tool to identify where data and resource needs exist. They could use information at a community level to identify potential partners and activities.

5. **Emergency and first responders** who are responsible for resource planning following an extreme

event. They can provide input about the type of information they need. They will use this information for grid recovery and to undertake appropriate resource planning efforts.

Stakeholders

Group L contributors identified stakeholders who must be engaged in the first phase of the OKN deployment are:

1. **Utilities:** Utilities are concerned about their own assets and spend significant amounts upgrading their infrastructures. Often this is not enough, as was seen during Hurricanes Laura (2020) and Ida (2022). Hurricane Laura left hundreds of thousands of people without power in Louisiana. Entergy (the utility company responsible for energy in Louisiana) spent millions upgrading their energy infrastructure, but the infrastructure failed again after Hurricane Ida, and caused wide-spread outages.

   Utilities and power grid operators can provide input on procedures and information they need to ensure the resiliency of the power grid. They can use the OKN to assess the risk to their infrastructures, potential impacts of these risks to communities that are served by specific utilities, and determine supply and demand relationship under normal vs extreme conditions and in case of renewables to address the justice and burden issues. This would be immensely helpful to smaller utilities with limited funds and rural user base, and they can use this information to identify their investment strategy under the Infrastructure Investment and Jobs Act.

2. **Local Communities:** Local communities often are unaware of their risk to extreme events when it comes to energy supply. They also don't have the full picture of the energy cost and the role of renewables in reducing the cost. This tool could be used by communities to assess the energy justice and burden issues in the case of climate extremes. This information could be used by utilities and communities to build public-private partnerships to reduce energy burden and greenhouse gas emission.

Human-centered design

Group L contributors identified three user groups and their needs and preferences for the Energy Management OKN. Needs and preferences will factor into the human-centered design for the OKN. Users and their requirements for the OKN are listed below.

1. Communities and community organizations (Southeast Energy Efficiency Alliance)
   a. Energy access and risk to energy infrastructures
   b. Risk and resilience profile of communities based on energy access
2. Emergency responders (FEMA)
   a. Risks to energy infrastructures for resource planning
3. Researchers (EPRI, Labs)
   a. Open datasets to evaluate the risk score for energy infrastructures

OKN Components — UI for communities, open source data access for emergency responders, ontology for researchers to assess risk and resilience profile.

Ethical and social implications

Given the use of public and private data, ethical considerations need to be addressed from the perspective of all stakeholders and users. The guiding principles for the OKN are:

- The overall data-ecosystem needs to conform to current rules, regulations, and FAIR (findable, accessible, interoperable, and reusable) practices
- Proposed use cases must adhere to the existing laws and rules (necessary but not sufficient)
  - Strengthen the existing rules and laws
  - Create oversight and accountability structure
- Consider whether the discretionary data access sharing/use should proceed — ethics analysis
  - OKN communities need to develop the ethics framework (consider 9 VA Ethics Principles for access to and use of Veteran Data)
  - Create oversight and accountability structure
- Consider a layered data model with various access rights
- Equitable access: intentional design, hinges on user needs, accessible and usable, actively engages underserved communities
- Equitable engagement (on data collection side)
- Community-based: From the design outset to data management
- Ethical and risk management frameworks for biases/assumptions
- Data Provenance to reliably identify the datasets and their relevance
- Protocols for data access and usage: Specific to types and sources of data
- Protocols and system in place to refute claims: Document disagreements
- Monitoring how data are used: Ethical use of user data

Path to success

The path to success for the Emergency Management OKN is to connect it with use cases developed by other user groups that participated in the sprint. Connecting with groups whose use case focus aligns with the Energy Management OKN would be of mutual benefit to the OKNs of all groups charged with creating separate but related use cases. Group L contributors recommend that Use Case L connect the use cases of the following groups:

- Group A (Strategic Supply Chain Resiliency)
- Group G (Climate Action)
- Group P (Compounding Disasters), and
- Group J (Decision Support for Government Leaders).

The needs of each group as they pertain to interconnection are listed below.

Group A: Access to materials for the production of decarbonization technologies is essential for

achieving energy grid and community energy resilience in the U.S. The lack of manufacturing as well as availability of certain materials within the U.S. border requires understanding the current energy supply-chain resilience and potential bottleneck.

Group G: Climatic events are one of the major drivers adversely impacting the current energy resilience. Often the lack of information about future climatic conditions as well as climate actions undermines the efforts of utilities and communities in achieving their energy resilience.

Group P: While the majority of energy resilience analysis undertaken by researchers and practitioners (utilities) focus on single events, the grid is at a risk of being impacted by more than one event. During COVID-19, the shift in population centroid impacted the energy supply-demand landscape by burdening the grid in certain areas. The occurrence of wildfires, hurricanes (Ida, Laura) and winter storm (Uri) also increased the burden on the energy grid to meet energy demand. The occurrence of multiple events and certain time-compounding events requires assessment of energy resilience in these conditions rather than in single-event conditions.

Group J: Multiple users, stakeholders and datasets pertaining to multiple systems (energy infrastructure, other critical assets, built environment, climate, social systems/population, environmental conditions) as well as knowledge about current policies are needed to determine the energy grid and community energy resilience. Furthermore, such analysis also requires to be conducted at multiple geospatial and temporal scales (neighborhoods, states, utility service areas) to assist with appropriate policy decisions. Often government leaders do not have access to all these datasets and information needed for decision-making.

The EMOKN (Energy management OKN) will connect with the afore-mentioned use cases to develop a holistic OKN that would allow exploring community energy resilience from the perspective of climates and compounding disasters, supply chain, policies and decision-making.

Sustainability and scalability

In the short term (1-3 years), the scalability component of the OKN should be addressed by establishing the connections with other OKN use cases as well as operationalizing the OKN discussed in this instance. From a sustainability perspective, the OKN should be expanded to answer other use cases and be used by users and stakeholders beyond those identified for this prototype. The OKN should also be adaptable to include new datasets and technological advancements and incorporate advanced visualization components. Partnership with private entities would also allow establishing the OKN as a business case to address energy resilience questions in different contexts and geographic scales.

Other concepts & issues

Group L contributors identified the establishment of a governance model to increase its usability as the one aspect crucial to the success of the OKN. The proposed governance model includes the following characteristics:

- Centralized governance at highest level possible (OSD ideally) to approve/issue guidance on overall taxonomy, requirements, etc. for data integration
  - Each service org will oversee day-to-day governance for their respective org (can be in the next version)
  - Data catalog and business glossary
- Hierarchical governance (sub-committees) to address:
  - Equitable access from all stakeholders
  - Data quality and provenance
  - Access to knowledge graphs for application development and domain mapping
  - Best practices for advancements

Quality assurance

Group L contributors answered the use case template question, "How will you know it's working as intended?" with careful consideration. They concluded that the best answer to this question is to use quality assurance measures to gauge the performance of the OKN. Quality assurance measures include:

1. Measure the speed at which new data is added
2. Time spend to reach consensus
3. Uptime of the overall infrastructure
4. Determine the time needed to standardize the data for integration
5. Conformance testing based on open standards and testbeds (similar to OGC)

Group L contributors will also review and adopt governance structures from large companies to address protocols for roles, communication and data access, and develop best practices to address multi-level governance structures.

Group M: Decarceration and Reunification
From Prison Incarceration to Community Reintegration and Family Reunification

The Decarceration and Reunification OKN is proposed to address the compelling need for more effective methods for reducing criminal justice system bias and providing fairer treatment and more support for incarcerated and decarcerated persons and their families. This OKN will also enable criminal justice and social services personnel to find new and better approaches to dealing with this population equitably.

Contributors

Contributors from government, academia, and the private sector volunteered to develop a use case for a Decarceration and Unification OKN. All contributors have knowledge and interest in this area, especially in reducing incarceration and recidivism, and removing bias from the criminal justice system.

Contributors' names and organizations are listed below.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Brian Handspicker	Open City Labs
Eric Jahn	Alexandria Consulting
Christina Medlin	Georgia Tech Research Institute
Sonia Kim	Alexandria Consulting
Daniel Stein	stewardsofchange.org
Kristine McCoy	stanfordalumni.org
Kenneth Berkowitz	U.S. Department of Veterans Affairs
Katherine Escobar	U.S. Department of Defense
Stephen Sullivan	U.S. Department of Defense

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Thomas Carlson	Georgia Tech Research Institute
Jennifer Hall	State of Colorado OIT
David Hardy	Booz Allen Hamilton

Champion

Brian Handspicker of Project Unify was the champion for Group M. As champion, Mr. Handspicker organized and led Group M meetings.

Reintegrating decarcerated persons reduces homelessness

The Decarceration and Reunification OKN would serve the public good by giving easy access to end-users and stakeholders from social services and other interested parties.. They would use the collective data obtained from the OKN to identify the decarcerated individual's needs for support and to determine his or her eligibility, or their families," to receive them.

Group M contributors propose building an Incarceration to Community Decarceration Service Planning system to support identification and delivery of social services. Incarcerated users need these services to enable their successful return to community and family life. The system would help achieve reduced recidivism and improved well-being. Both the formerly incarcerated individuals and their family members would benefit. **Figure 10** depicts the system.

Group M contributors incorporated the Family Unification Program Housing Choice Voucher (HCV) as part of decarceration service planning. **Figure 11** depicts steps in obtaining the continuum of care that decarcerated individuals and their families need. HVC can improve a family's ability to attain/retain stable housing and enable the return of children to their families from out-of-home care after an incarcerated parent's return to their community.

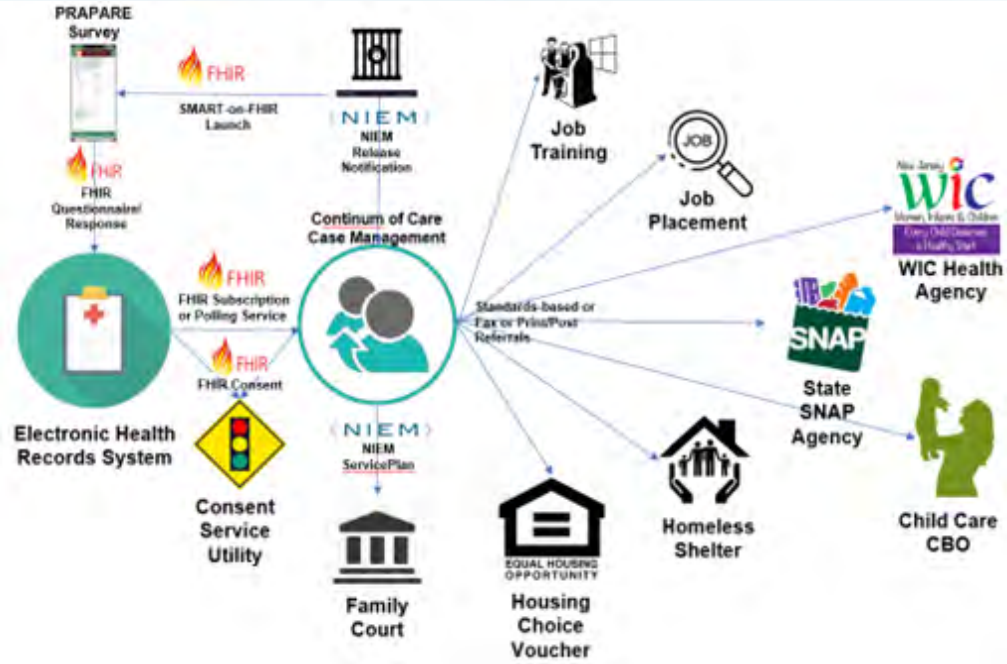


Figure 10: Decarceration Overview Dataflow

Serving the public good

The proposed Decarceration and Reintegration OKN serves the public good by providing integrated, shared, and easily accessed information that social services can use to support the successful reintegration of incarcerated individuals into society. Finding housing, healthcare, and food and financial support enable a decarerated person to resume life as a free citizen.

His or her successful reentry into society helps reduce crime and homelessness, both of which add to the financial burden of communities that must pay for more policing and declining community health. Crime increases when people without housing and resources attempt to sustain themselves. Incarceration depletes community treasuries used to hire additional police officers and prison guards. Community health suffers when decarcerated individuals have no option but to live on the streets.

Goals and objectives

The goal for the Decarceraction and Reunification OKN is to build an OKN that will help incarcerated individuals successfully return to community and family life, resulting in improved the well-being of the individual and their family.

By successfully reintegrating with their community and reintegrating with their family, recidivism will be reduced and community well-being will be improved.

This use case incorporates multiple different systems, each with distinct domain-specific ontologies. This means that potential users will use their native systems to access the resulting broader OKN. This tool will be used by:

- Prison Administration and Prison Social Workers
- Community Continuum of Care Managers
- Community Social Services Managers
- Community-based Health and Social Service Providers
- Homeless Shelters
- Child Welfare
- Housing Providers

Achieving objectives

To achieve the objectives of reintegrating a decarcerated individual into society and reunifying him or her with their family, the scope of services must be connected to needs. The OKN must accommodate this. Identifying these needs in context requires identifying multiple threads pertaining to the services that the decarcerated individual may require. These threads are listed below.

1. Multiple threads/use cases of a common user story:
    a. Justice thread — transparency of arrest, prosecution, conviction, imprisonment, rehabilitation, release planning, release, parole, post-parole outcomes
    b. Individual Health and Human Services thread — pre-release needs assessment, referral to health and social services, socialcare assessment, service plan/care plan, service referrals, parole officer on service/care team
    c. Family Human Services thread — (b) above plus Continuity of Care engagement, Family Housing Voucher, child welfare and family court added to service/care team, foster care plan

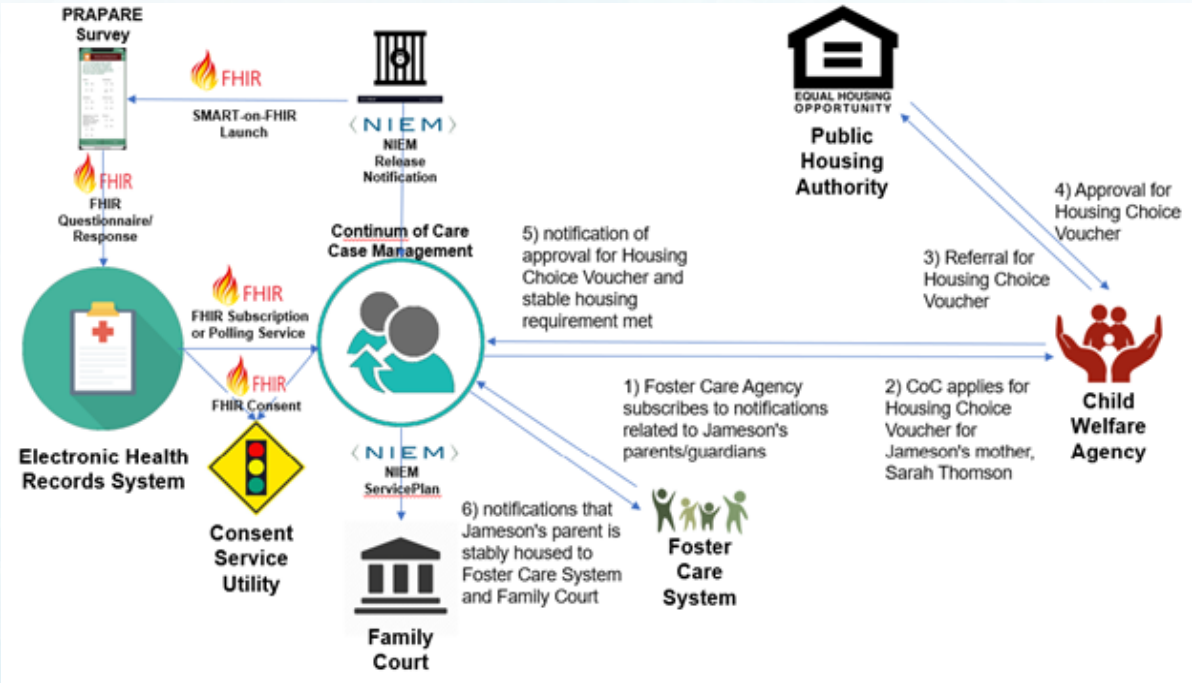


Figure 11: Reunification Housing Choice Voucher Dataflow

Challenges

Group M contributors identified the following challenges to building a Decarceration and Reunification and in maintaining its integrity after it is built. Challenges to building the OKN include:

- Lack of Ontologies for NIEM Domains
  - Stand up NIEM Ontology for NIEM Core, NIEM Justice (see NIEM Ontology use case)
- Lack of Ontologies/Data Standards for Human Services
  - Identify Use Case Target Services,
  - select associated Domain models
  - Define Ontologies
- Lack of Interoperability Standards between Domains
  - Identify Use Case Domains,
  - select associated Domain standards
  - Map between Domain-specific ontologies and interoperability standards
- Lack of Common Consent Services
  - Build Common Consent Service (see Consent use case)

Challenges to maintaining integrity in serving the public include protecting privacy, improving integration within the community, and enabling family reunification. Meeting these challenges includes:

- Offering public/personal control over privacy preferences and consent
- Enabling improved integration of decarcerated individuals into their community
- Enabling reunification of decarcerated individuals with their families

Meeting these challenges will help to maintain the integrity of the OKN and thus gain the public's trust. It will also result in lower recidivism, safer communities, and increased contributions by individuals to their communities and families.

Resources

Group M contributors identified existing resources that could be incorporated into the Decarceration and Reunification OKN.

Existing resources

Existing resources useful if not essential to OKN development include:

- Ontologies/Data Models, including:
  - Justice (NIEM Data Model—- need Ontology)
  - Healthcare (FHIR)
  - Continuum of Care (HUD HMIS)
  - Homeless Management (HUD HMIS)

- Ontologies for:
  - NIEM Core and Justice Domains
  - Case Management
  - Child Welfare
  - Housing
  - Food Social Services (SNAP, WIC, etc.)
  - Transportation
  - Employment training and counseling

- Required Services and Tools
  - Consent Services
  - Ontology Definition and Mapping Tools

Other expertise

Group M contributors identified the following areas of expertise needed for OKN development. They also identified certain individuals whose service would be valuable in this effort. Other areas of expertise include:

- Subject Matter Expert (SME) with domain expertise in the following areas:

  - Prison Social Work and Administration
  - Criminal Justice System
  - Juvenile Justice (e.g. in user story of incarcerated teenage single mother)
  - Healthcare Coordination — Dr. Kristine McCoy
  - Continuum of Care Coordination
  - Homeless Management — Eric Jahn
  - Socialcare Case Management — Brian Handspicker
  - Child Welfare
  - Housing
  - Food Social Services (SNAP, WIC, etc.)
  - Transportation
  - Employment training, career counseling, job placement
  - Civil legal aid

End-users & stakeholders

Group M contributors answered the following use-case template questions to explain how Decarceration and Reunification OKN will serve the public good. They identified standards and OKN use cases that would support OKN development and ensure its success. They also identified end-users whose user requirements are essential to OKN development and use.

1. How will your OKN be utilized?

This OKN will be utilized to enable Community Decarceration Service Planning. The OKN will support identification and social services delivery. Social services enable the decarcerated individual's successful return to community and family life.

2. Serve the public good?

This Knowledge Network and associated case management and service provisioning would serve the public good by reducing recidivism and improving personal and community well-being.

3. By which users?

It will be used by incarcerated individuals and the following types of agencies/organizations engaged in the creation, maintenance, and delivery of the individual's Decarceration Service Plan. Agencies include:

- Prison Administration and Prison Social Workers
- Community Continuum of Care Managers
- Community Social Services Managers
- Community-based Social Service Providers
- Homeless Shelters
- Child Welfare
- Housing Providers

Applicable standards

Group M contributors identified international and national standards that must be included in the development of the Decarceration and Reunification OKN. These standards include:

- HL7 Community-Based Care Privacy WG (Consent Services, Data Segmentation for Privacy)
- HL7 Human and Social Services WG (Consent for non-health)
- HIMSS PP2PI WG (Data Segmentation for Privacy requirements)
- San Diego Health Connect (CDS-LEAP implementer)
- National Interoperability Collaborative Project Unify WGs
- National Information Exchange Model PMO

OKN use cases

Group M contributors identified the following OKN use cases as essential to Decarceration and Reunification OKN design and build.

- NIEM Ontology
- Consent Service
- Homeless Tracking
- Court Judicial Event Tracking
- Improving Transportation for Health

International, national, and local end-users

End-users whose requirements for the UI and user controls for the OKN are personnel in the government agencies where they work. These end-users and their organizations would receive significant benefit from the development and deployment of the OKN. They include:

- Prison Social Work and Administration
- Criminal Justice System
- Juvenile Justice (e.g. Decarceration user story of incarcerated teenage single mother)
- Healthcare Coordination
- Continuum of Care Coordination
- Homeless Management
- Socialcare Case Management
- Child Welfare
- Housing
- Food Social Services (SNAP, WIC, etc.)
- Transportation
- Employment training/counseling, job placement

Human-centered design

Group M contributors interviewed end-users and stakeholders to identify what they want and need from the Decarceration and Reunification OKN. Interview results, edited for concision and clarity, become requirements in human-centered design. HCD user requirements will be included in OKN development. They are essential to the usability of the OKN and to its success.

End-users interviewed for their requirements represent the government agencies where they work. The requirements gathered document the wants and needs of these users, who will likely use this OKN daily in their work. Interview results for each government agency are listed beneath the government agencies that are the end-users' employers.

End-users

1. Justice System and Prison Administration:
   The justice system in general, and prison administration specifically, needs tools to manage decarceration service planning and tracking the follow-through on those plans once an individual has returned to their community and family.

2. Health and Human Service Providers:
   Service providers need a semantically represented data and associated consent framework for cross-domain consent presentation and retrieval, enabling service record sharing, eligibility determination, and client matching.

3. Analysts from public safety, public health, and human services:

Analysts need to be able to track end-to-end information about an individual in the criminal justice system during adjudication, during incarceration, and beyond release back to community to determine if proactive access to health and human services reduce recidivism and improve individual, family, and community well-being.

Stakeholders

Stakeholders represent the agencies listed below. The results of their interviews are displayed beneath each agency heading.

1. NIEM Ontology:
   The NIEM Ontology proposal needs a use case to demonstrate the exchange of NIEM Justice Domain data with non-NIEM domains.
2. Homeless Continuums of Care:
   Continuums of Care need a way to manage arrival of potential shelter/housing residents as they are released from prison to a community. They also need to be involved in service planning for other human, social, and health services.
3. Justice Tracking OKN:
   This Decarceration/Reunification OKN is potentially an ideal use case for the Justice Tracking OKN proposal.

Ethics & social implications

The ethical implications of Decarceration and Reunification OKN use pertain to the potentially negative consequences of sharing information.

Beneficence is a guiding ethical concept of security, privacy and consent based information exchange. In the absence of mechanisms, policies and governance processes that support definition of and enforcement of privacy preferences (requirements) and consent, there is significant risk of causing unanticipated or unintentional harm.

To mitigate the possibility of negative consequences, individuals whose information will be shared should be able not only to provide consent for access to information at a point in time, but also to:

- Review the record of their data as it is exchanged through various phases and be permitted to revoke consent at any time.
- Review clear documentation of the ways their data have been coded and integrated with other data.
- Object to classifications, logged records and specific uses of their data.
- Request changes to and deletion of any inaccurate information.
- Be alerted to the possibility of false positives or negatives in identity matching, and be able to review and contest such matches. This should also apply to the other person(s) whose identity is being confused, conflated or mistakenly matched in the process.
- Monitor and comprehend the ways (for institutional users) in which aggregate data is used in

algorithmic decision-making processes and predictive analytics.

Understanding stigma and other sensitivities

Another ethical implication of Decarceration and Reunification OKN use is that users may not fully understand the care to be taken with the potentially sensitive information that can be accessed from the OKN.

To the extent that open knowledge network information integration includes potentially sensitive individual, family, or organizational information, consent issues should be understood in a context of stigma which may drive disincentives for self-reporting information, effectiveness of use of information, and the risks of information being used by non-consented systems. Other specific sensitivities include:

- Ongoing discussions in which statements like "replace the child welfare system" rather than reform it, are used can have negative consequences. Citing "over-surveillance" of racial groups as a diversity/equity issue that results in the disproportionate removal of Black and other minority children is a negative consequence.
- The deep stigma attached to drug-using parents creates a significant disincentive for parents to self-report or consent to reports. Laws in half the states define prenatal substance use as formal child abuse reportable to child welfare. Drug-using parents do not self-report for fear of child removal and/or incarceration.
- Inconsistent judicial rulings regarding persons with SUDs who are complying with prescribed medication-assisted treatment, but who are assessed as non-abstinent and thus non-compliant.
- The publicity given to recent breaches in data-system security, leading to hundreds of thousands of records becoming available to unauthorized users and ransom-seekers.

Information-sharing ethics

Data Sources	Data Policies	Data Rights	Data Purpose	Data Limitations
Prison Administration	Privacy Act of 1974	Consent, Privacy Prefs	Plan Reentry	Impact Employment
Healthcare	HIPAA	Consent, Privacy Prefs	Reentry Health	
Continuum of Care	HIPAA	Consent, Privacy Prefs	Coordination	Stigma
Homeless Shelter	Homeless Bill of Rights	Consent, Privacy Prefs	Initial Housing	Stigma
socialcare Case Mgt	HIPAA	Consent, Privacy Prefs	Coordination	
Child Welfare	COPPA	Consent, Privacy Prefs	Reunification	
Housing	Privacy Act, FHA	Consent, Privacy Prefs	Housing	Stigma
Food Social Services	Privacy Act of 1974	Consent, Privacy Prefs	Food	Stigma
Transportation	Privacy Act of 1974	Consent, Privacy Prefs	Transportation	
Employment	Privacy Act of 1974	Consent, Privacy Prefs	Employment	

Table 1: Essential elements of ethical standards for information sharing

Table 1 displays the elements of ethical standards for sharing information. Ethical standards must be adhered to in the handling of this data, both in OKN development and on the job. Adherence ensures that service providers understand data policies and follow them accordingly. Service providers and other users breach ethical standards when they fail to learn and understand ethical requirements for information sharing.

Information-sharing consequences

Misused, these data sources, policies, rights, and purpose could have negative consequences for clients and service providers.

To prevent misuse, the following questions should be answered

- Do people understand the purpose?
  - Yes, in general, though there are always some that need positive education.
- Who will be positively affected?
  - Incarcerated individuals
  - Families of incarcerated individuals
  - Communities into which incarcerated individuals are released
  - Community socialcare service organizations (CBOs)
- Who could be negatively affected?
  - Incarcerated individuals and/or families whose information is misused
- How are we minimizing negative impact?
  - Enhanced privacy preferences and consent services
- How can people engage with the service?
  - Consent services, privacy preferences, Information change requests, appeals

Path to success

The path to success for the Decarceration and Unification use case is to understand its purpose, uses, dependencies, and risks. Understanding these parameters to use-case development helps ensure that the use case will be developed correctly. The following questions elicit information useful in use-case development.

Questions for use Case development

Questions asked to identify optimal use case connections yielded the following results, including optime use case connections, use case dependencies, and use case associations.

1. The use should integrate as a Community Care Information and Resource Matching with:
   a. Group M: Decarceration/Reunification to ensure planned socialcare investment in reintegrating decarcerated persons with community and reintegration with family
   b. Group B: Integrated Justice Platform tracking outcomes due to socialcare investment in reintegration into community/reunification of families
   c. Group H: Homeless Tracking information since often the first night out of incarceration may be in

a homeless shelter
   d. Group P - Compounding Disasters/Events (AKA Wicked Problems Institute) which needs to provide socialcare to displaced persons/families
2. This use case is dependent on:
   a. D: NIEM Ontology
   b. I: Electronic Consent
   c. Ontology Development and Data Modeling Mapping Tools
3. This use case may be associated with:
   a. B: CJET for tracking of outcomes due to human services investment in reintegration into community/reunification of families
   b. H: Real-time Homeless Information since often the first night out of incarceration may be in a homeless shelter
   c. Improving Transportation for Health since reintegration in a community, access to jobs, and follow-up healthcare will be dependent on access to transportation

Sustainability and scalability

Group M contributors envisioned sustainability and scalability for the short-term (1 - 3) years and the long term (4 - 15) years. Their proposed activities for short-term and long-term time frames are listed below.

- Short-term: invest in the creation of Ontologies/Data Models for socialcare domains that are needed to support Decarceration and Reunification
- Short-term invest in definition of interoperability standards for socialcare domains to enable creation of OKNs across multiple socialcare domains
- Short-term: support pilots of Decarceration and Reunification in target communities
- Long-term: encourage and support the use of Decarceration and Reunification OKNs across many communities

Other concepts & issues

Group M contributors identified the following additional concepts and issues crucial to use case development:

- Identify/define privacy and consent regulations, policies and rules for socialcare domains
- Extending/constraining/leveraging standards for privacy and consent for socialcare domains

Group N: The Education Gateway

Group N aims to develop an OKN Education Gateway which will provide a unified learning platform for OKN architects, data contributors, researchers, as well as for end users and stakeholders. Contributors represent government agencies, academia, and business.

Contributors

Group N contributors' names and organizations are listed below.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Bandana Kar	Oak Ridge National Laboratory
Emanuel Sallinger	Vienna University of Technology and Oxford University
Emily Rothenberg	Columbia University
Lauren Close	Columbia University
Ora Lassila	Amazon.com
Nariman Amma	The University of Tennessee Health Science Center

Champions

Group N champions organized meetings and led discussions for Use Case N development. Their names and organizations are listed below.

CHAMPION'S NAME	CHAMPION'S ORGANIZATION
Cogan Shimizu	Kansas State University
Esther Jackson	Columbia University
Florence Hudson	Columbia University

The Education Gateway:

An Entry Point to All OKN Knowledge Networks

This OKN Use Case will act as an entry point to the rest of the OKN, by providing a unified, foundational learning platform. The different aspects and initiatives of the OKN will use a wide variety of different tools and technologies, many of which will leverage fundamental concepts that are likely new to many of those involved. These concepts, the tools that are built on top of them, the frameworks they employ or comprise; and the use cases that they solve will all need to be described in an understandable and approachable manner, so that the OKN may best serve the public good and facilitate true public access.

This OKN Education Gateway will provide learning and outreach materials for...

A) Agency administrators, elected officials, and other stakeholders who need to understand what an OKN is, why they are important, how they can be leveraged, and the support they require (e.g., compute, data, and funding).

B) OKN architects, data contributors, professionals, educators, researchers, and students who need to ensure convergence of the technical foundations, drive data integration, identify gaps in the open knowledge, contribute to the OKN, and train the next generation of Open Knowledge Engineers.

C) Members of the public and community scientists who need to understand how to use, interpret, and contribute to the OKN.

Each role in the OKN ecosystem will be treated as a first-class citizen, not as an afterthought, with dedicated curricula and materials, to ensure each role learns what is applicable to them.

Resources

In order to enact our vision,
...we will utilize resources that do exist:

- **Learner Central**, **Educator Central**, **Video Library** (National Student Data Corps)
- **Data Science Resource Repository** (NE Big Data Innovation Hub/NSDC)
- **The Carpentries** (for instructor-led teaching model that is remote-friendly)
- **How to Create a Knowledge Graph?** (Stanford)
- **Knowledge Graphs** (arXiv)
- **Ontology Wikipedia Article**
- **What is AI?** (Stanford)
- **Knowledge Graphs and Big Data Processing** (Oxford/TU Wien/LAMBDA)
- **Explained: Neural networks** (Deep Learning, MIT)
- **Machine Learning, Explained** (MIT)
- **The Missing Semester** (CS education, MIT)
- **Designing and Building Enterprise Knowledge Graphs** (book)

- W3C standards documents (rdf, sparql, owl, etc.)
- REAL Volunteers (NEBD)

...we will need to develop resources that do not already exist:

- A web portal for self-paced, curriculum-based learning on any of these topics
  - Stage 1: Microvideo library for specific, yet general or foundational, topics
  - Stage 2: Learning materials that are explicitly domain-specific
  - Stage 3: Technology modules — GraphPlaygrounds and Code Sandboxes
- A Support Network for new and experienced users/OKN users; it is important to include subject specialists from different government agencies who have data in OKN
- Robust User Profiles
  - Stakeholders: data contributors, policy makers, elected officials, funders
  - The Curious Public: Laypersons to Experts
  - Educators and Learners
  - Developers and Implementers
- For each User Profile include "Inspirational" and "Aspirational" Material
  - Value Propositions: inspiration for using the technology in new and creative ways
  - Effective Onboarding: learning paths for newcomers to the Education Gateway
  - Customized Perspectives
    - Simplified Examples: narrow use-case scenarios
    - Wide Examples: show how KG and OKNs can serve as a platform
    - OKN Management: governance, evolution, project guidance and leadership

...we will need to build out a robust team covering a wide variety of competencies:

- News & Blog Writer for Press and Outreach
  - Academic Outreach
  - Government & Federal Outreach
  - Industrial & Private Outreach
  - Public & Community Outreach
- Community Driver
- Educational Design Lead
- Documentation Lead & Technical Writer(s)
- User Interaction & eXperience Lead
- Outcome Assessment Lead
  - Responsibilities: Identify Top-Level Goals and determine if they've been met
- Learning Assessment Lead
  - Responsibilities
    - Metrics for effectiveness of our learning materials
    - Badges and certificates
    - Some strategy that indicates that this is a closed loop model
    - REAL Volunteers (NEBD Hub & NSDC)

- Domain and Subject Matter Expert Contacts
  - People who can explain how the data in an "established OKN node" can be understood and utilized.

...we will need to be mindful of resource requirements that evolve as the OKN evolves:

- As the OKN is developed, the various OKN nodes may use different tools and techniques, entry points, identity and access management rules, and others which we will need to understand and incorporate
- New technical resources may be needed to work with the various OKN tools and techniques
- Educational instructors may be needed
- Additional support may be required for things like video lectures (e.g. hosting, DOI registration, and transcription services).
- New teaching materials may need to be developed (for different parts of the OKN, for domain-specific scenarios), in keeping with some agreed-upon standards

End-users and stakeholders

Group N was formed during the Innovation Sprint to acknowledge the critical lack of educational and learning material regarding OKNs and their related technologies. Our team represents interests from academia, government agencies, and industry. We have the knowledge and expertise to develop a highly visible and effective Education Gateway.

Group N contributors envision the Education Gateway as a learning platform for all users. We have identified end-users and stakeholders within and without the Innovation Sprint who have an immediate and clearly defined need for the Education Gateway.

Serving the public good

The Education Gateway will serve the public good by providing a broad range of users access to a broad range of information essential to finding solutions to the serious problems that the nation and the world now face. Researchers, policy-makers, and others with the ability to make change can build on existing knowledge pertaining to current and anticipated threats to the environment and public health and safety.

The Education Gateway will be utilized by any person in the public to learn about the content, domains and their interrelations; how to leverage the OKN; and the technologies that support the OKN. Facilitating OKN accessibility is fundamental to serving the public good, and overall supports each other OKN use-case in serving the public good. The gateway will minimally consist of learning material customized for various learning perspectives, and similarly contain guidelines for developing domain and OKN use-case specific curricula that will ensure a coherent and consistent presentation of OKN documentation, in turn facilitating greater public engagement with the OKN.

Engaging stakeholders

The Education Gateway is grown and sustained by both symbiotic relationships with the other OKN use-cases. We will also engage representatives of our broad, public demographics, such as graduate students and members of the curious public.

Human-centered design

For each group, we have conducted an interview (synchronously or asynchronously, depending on availability). End-users have been engaged via survey. In many ways our stakeholders are also our end-users. Any stakeholder would find use in the outcome, and any end-user should want this use-case to succeed; thus the lines are quite blurry.

Internal stakeholders

For our internal stakeholders, we reached out to several Groups that have an emphasis on transdisciplinary research and expect to interact with many different types of end users.

- Group D - NIEM Ontology (Eric Jahn)
- Group P - Compounding Disasters (Ryan McGranaghan)
- Group H - Homelessness Knowledge Network for Cities (Samuel Klein)
- Group Q - Public Risk Analysis & Alert System via Financial OKN (Oktie Hassanzadeh)
- Group F - An OKN for Health Communication Management (Chengkai Li)

External stakeholders:

- Amazon Web Services (Ora Lassila)
- The Knowledge Graph Conference (François Scharffe)

End-users:

- PhD Students (Data Semantics Laboratory @ Kansas State and Wright State Universities)

Ethics & social implications

Group N contributors envision the Education Gateway as a learning platform for all users. We have identified end-users and stakeholders within and without the Innovation Sprint who have an immediate and clearly defined need for the Education Gateway.

Limitations, potential biases, gaps & omissions

As educators, we must ensure that our data (i.e., our educational materials) do not contain implicit bias. This can be mitigated by providing sufficient foundational material to make the content accessible. From another perspective, examples can be closely verified to be inclusive. We will communicate closely with our learning community on the necessary learning objectives so that all needs are met from both an ethical and technical perspective.

Legal and ethical permissions for the reuse of data

All reused material, such as from educational partners (e.g., professors), will require legal permission. However, we foresee that much of the educational material will be bespoke and created by and for the Education Gateway.

Personally identifiable information.

The Education Gateway does not require any personally identifiable information. We do not intend to collect personally identifiable information.

Negatively impacted demographics

The Education Gateway must take into account the intersectionality of educational privilege across demographics, from both a socioeconomic perspective, and simply as a matter of expertise. In order to mitigate this, we intend to survey multiple audiences in order to foster a widely available and accessible Education Gateway to the OKN. We will include (and closely monitor) a feedback mechanism to identify gaps and possible negative impacts from the wider learning community.

Path to success

Identifying, curating, and developing new learning material is fundamental to the initial success of the Education Gateway. However, the true path to success is continuous adaptability as the OKN users' needs change. We discuss these key factors below.

Connectivity

The Education Gateway is not, in particular, an instantiation of the overall OKN. Instead, it is a fundamental aspect of any particular instantiation, or the proto-OKN. It supports every other use case, by providing educational materials to end-users, stakeholders, and other interested parties. End-Users and Stakeholders identifies several groups where education plays a prominent role.

Sustainability and scalability

In the short term, the educational materials and related curricula need to be created. In the long term, of course we would need to ensure that the curriculum remains up-to-date and speaks to the state of the art. Over the entirety of the lifespan of this project, we would want to ensure that we are adapting new educational vectors, developing instructional infrastructure for effective learning, and broadening participation in OKN technologies.

Group P: Compounding Disasters/Events
The Wicked Problems Institute

Use Case Group P aims to create a knowledge infrastructure that brings together existing OKN efforts that address specific risks such as urban flooding, forest fires, space weather and terrestrial weather storms and other natural disasters to understand the compounding effects of risks, and to create situational awareness for any given disaster management task.

Contributors

Group P contributors are listed below, along with their organization.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Ashley Atkins	University of California, Berkeley
Bandana Kar	Oak Ridge National Laboratory
Christine Kirkpatrick	San Diego Supercomputer Center
Dalia Varanka	U.S Geological Survey
Ellie Young	Common Action
Florence Hudson	Columbia University
Ilya Zaslavsky	University of California San Diego
Jeff Heflin	Lehigh University
Johannes Friedrich	World Resources Institute
Kenia Pujols	Columbia University
Krzysztof Janowicz	University of California Santa Barbara
Lauren Close	Columbia University

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Lilit Yeghiazarian	University of Cincinnati
Michel Biezunski	Infoloom
Sharat Israni	University of California San Francisco
Silviu Cucerzan	Microsoft
Cogan Shimizu	Wright State University

Champion

Ryan McGranaghan from Atmospheric and Space Technology Research Associates (ASTRA) LLC was the champion for Group P. As champion, Dr. McGranaghan organized use-case group meetings and briefed contributors on group updates.

The big picture

Group P aims to create an information commons that brings together existing OKN efforts that address specific risks (e.g., floods (**UFOKN**), fires (**KnowWhereGraph**), space weather and terrestrial weather storms (**CHESS**) and integrative tools (the **COVID Information Commons**, **EarthCube** and **Big Data Hubs**) to understand the compounding effects of risks, and to create situational awareness for any given risk.

Few examples of how Use Case P will serve the public good are listed below:

a) A 'day in the life' - power grid engineer responding to conditions on the grid they monitor at any time using the OKN. As residents in Texas and elsewhere around the world can attest, the forces of the natural world together with the vicissitudes of human behavior have the power to wreak havoc on the power grid. Power grid responses are largely reactive (not proactive), formulaic, and under-informed. Those creating the policies that determine these responses do not have seamless access to the cross-domain data/information to guide more informed decisions driven by a more systemic/complex understanding.
b) The challenges of today require complex data integration and novel sources of information to make rapid decisions. Disasters are layered and cascading, from wildfires that cause landslides that cause economic, safety/transportation hardships as well as how pandemics change old models to disaster response.

c) Group P information commons will create situational awareness for risks as complex interconnected events

Examples of potential users of the proposed tool include

a) Researchers to develop models to increase situational awareness
   i) **Contextual, adaptive, probabilistic and predictive** when possible
   ii) Uncertainty quantification for guidance provided to any event/hazard/risk
b) Policy-makers/Action-organizations - the goal is to **get more *proximate* to the community**
   i) Infrastructure managers (e.g., power grid operators)
   ii) Government leaders and agencies
   iii) Citizens/residents, visitors, caregivers, enterprise leaders in the area.
   Example: https://cusec.org/
c) NOTE: The system will serve different users depending on the phase of the risk that they are addressing: pre-, during-, and post-disaster times

The grand narrative

The number and severity of disaster and disruption events is expected to increase in communities across the United States. Disaster events affect multiple community infrastructure systems (e.g., energy, supply chain, built infrastructure) simultaneously, producing a range of dynamic disruption situations. Disaster events also occur within a continuum of other events, such as disease outbreaks. The co-occurrence of more than one disruption event, such as a cholera outbreak together with a hurricane that damages sanitation systems, results in a more complex impact scenario, which in turn requires a customized management response.

In any disaster context, restoring stability to individuals and systems requires coherent response from community members, disaster responders, and infrastructure managers. Responses must be quick to respond to threats within evolving and uncertain contexts. Thus, disaster response participants benefit from access to crucial decision-support information presented in an understandable, interactive, and rapid manner. However, supplying such information is a pernicious challenge, as information needs are so complex. For instance:

- **Professional responders require unpredictable combinations** of technical information spanning disparate domains, which is typically not well integrated.
- **Operators of related systems (e.g., power grid, police, transportation departments, etc.) must align response decisions** within broader systems contexts, which may be influenced by factors of which they are not aware. Further complicating operator response is the need to follow predetermined procedures that may be brittle to hazard or disaster contexts.
- **Each event requires a unique assessment of local damage and other related phenomena** (e.g., the status of community sanitation systems), as well as specific information about individuals present within the disaster setting (e.g., health risks, medication needs, age). This information is required at various temporal horizons; for instance, logisticians require information about local

community needs when determining what type and number of supplies to deliver, a decision that precedes local arrival.

- **Response efforts require coordination between a dynamic range of actors who must quickly identify and establish working relationships**, and manually decide how to collaborate to achieve goals. Ideal information flow requires horizontal (e.g., peer), vertical (e.g., hierarchical), and cross-network transmission, yet there exists no support mechanism to deliver this. The central challenge is information sharing that builds trust, i.e., strong networks.
- **Individual community members must respond to needs and risks for themselves, families, and proximate others**, which requires safe navigation of unstable physical environments and up-to-date information about transport systems, community support opportunities, location of healthcare, food and water, and access to various other resources (e.g., tools, vehicles, etc.)
- **The culmination of the previous two bullet items are that responses must be multi-level** (across individuals, families, small communities, and up to societies).

An OKN, as a platform that enables the aggregation and connection of diverse and multi-scale data, can address these information problems. An OKN could provide a technological framework for delivering "just-in-time" information to decision-makers operating at every layer of the response effort, enabling:

- Community members to identify pathways and resources for meeting emergent needs
- Professional disaster responders to access historical situational information that can help contextualize and real-time situational information, to identify the full set of factors present in the real-time disruption context
- Professional systems operators to make informed decisions within systemic contexts
- Convergent representation of the common problem theater, through citizen-reported information reporting, enabled by an app like **Ushahidi**
- Efficient knowledge transmission between users across the network, eliminating bottlenecks and gaps in communication lines

To develop an OKN in this context, Group P endeavor to identify specific user-centric use cases to serve as starting points for identifying further contextual information about existing needs and questions, along all axes of disruption:

1) Community needs, such as electricity, food, water, shelter, communications, sanitation and others identified by groups such as **Resilience Maps**
2) Community infrastructure that typically meets those needs (supply chains, energy systems, etc.) and available alternatives
3) Predictive disruption potential for such systems, enabling proactive responses in community systems, such as the power grid
   a) *Use case: "A system that allows a user to overlay effects due to various hazards on an infrastructure that they are interested in (e.g., space weather + wildfire impacts on the power grid); map-based valuable"*
4) Anticipation of compound failure potentials
   a) *For additional details see* **Dong et al**.

5) Identifying existing questions that communities face when responding to disasters, to identify points at which information can be usefully delivered
6) Identifying existing gaps in information supply for disaster responders and managers of local systems, through dialogue with organizations such as the American Red Cross, FEMA, AHCU.S.A, etc.

Resources

Group P categorized their resources into existing resources, unknown resources and evolving resources. More details are given below:

Existing resources:

Communications Infrastructure

- Existing emergency response apps (RedCross, FEMA forms, etc.)
- Local communities (**Ushahidi**)
- Real-time information share between disaster responders

Existing schema/ontologies for hazards/disasters

- Ontology: An ontology for space science hazards that could link to Earth Science hazards: **https://arxiv.org/abs/2009.12285**
- WIFIRE data commons and ontology
- COVID Information Commons (**covidinfocommons.net**) [Florence]

Datasets on specific disaster

- Space Weather geomagnetic indices (proxies for space weather activity)
- Geomagnetically Induced Currents in the power grid (data available for historic space weather storms)
- Ground-based magnetometers (SuperMAG) that quantify geomagnetic disturbance
- Emergency preparedness documents at different government levels
- NHERI DesignSAFE and Natural Hazards Engineering Community **www.designsafe-ci.org/** and Multi-Hazard Engineering Collaboratory on Hybrid Simulation (**mechs.designsafe-ci.org/**) [Florence]

User communities:

- User Community: Sun-to-Power Grid **https://www.chessscience.com/sim-game-workshop**
- User Community: NASA Space Science
- Use Case: Terrestrial and Space Weather impacts on the Power Grid: **https://www.chessscience.com/**

Compute/infrastructure resources

- KnowWhereGraph, Urban Flooding OKN
- From EarthCube
- Gleaner
- GeoCodes
- Throughput

- Council of Data Facilities (geoscience repositories)
- Research Data Hubs
- COVID Information Commons
- **COVID-19 Community**
- **The Natural Hazards Engineering Research Infrastructure**
- OpenStreetMap
- World Resources Institute - Has an assessment of natural hazard mapping tools (Johannes Friedrich)

Resource unknowns:

- What are the combinations of datasets needed to respond to individual disruption events?
- What are the evolving needs during disaster events?
- How can real-time information sharing be adequately supported?

Evolving resources:

- The usefulness of the OKN will improve over time, as datasets are developed iteratively, in response to several real world circumstances

End-users and stakeholders

End-users and stakeholders are organized into Internal (within the Wicked Problems Use Case) and External (within other connected use cases). Internally, Group P categorize users and stakeholders by phase of the hazard:

- **Pre-** (e.g., Decision makers to choose investment priorities for disaster resiliency and Organizations evaluating where to build a facility);
- **During** (e.g., First responders need to identify how the situation/risk is changing and what areas are impact or at risk and Citizens); and
- **Post-Disaster** (Decision makers to prioritize repair actions during the recovery phase; Community to know how to stay safe during the disaster event, and to be informed about what actions are being taken to restore their services and how long restoration will take

Group P has membership distributed across various OKN innovation sprint use cases and will enable us to factor in their user information

Horizontal

- Group N: Education Gateway
- Group D: The NIEM Ontology

Vertical

- Group G: (community climate action)
- Group E: (nutrition to climate)
- Group L (energy resilience)

Stakeholders by topic

Group P has (or will) engaged the following users/stakeholders

- **Space weather** (note interviews and engagement with each of the following have been carried out by a recent three-day workshop **https://www.chessscience.com/sim-game-workshop)**
  a. Power grid operators and managers (more than ten utilities across North America)
  b. Policy-makers (FEMA, FERC, NERC, Space Weather Advisory Group)
  c. Researchers (NASA, NOAA Space Weather Prediction Center, NSF Geospace Environment Modeling and Space Weather)
  d. Future targets: utilities worldwide; citizens affected by power outages; Climate and Terrestrial Weather Scientists (understand the compounding effects)
- **Fire**
  a. The General Public: Evacuation management and predictions, logistics and transportation interfered by fire and smoke plume locations
  b. Consumers: Crops contaminated by ashfall,
  c. Researchers: Causal Event analysis (e.g., landslides caused by foliage burnout)
  d. Cyberinfrastructure providers, platforms to integrate with or who will use the OKN (e.g., **WIFIRE**)
- **Flood**
  a. Watershed management stakeholders (e.g., sewerage and drinking water utilities) with community stakeholders (via Group G)
  b. Municipalities, state agencies, especially with history of flood events
  c. Federal government agencies and nonprofits, especially those directly supporting flood recovery (e.g., FEMA, Red Cross)
- **Climate**
  a. Energy companies/utilities (via Group L)
  b. Insurance companies/actuaries
  c. Researchers (NSF, NIH, NASA funded)
- **Health**
  a. Federal, state and county level public health agencies
  b. Researchers (NIH, NSF and DoD funded)

Ethics & social implications

As part of ethics and social Implications, Group P identified the following four topics as the main areas of concern:

Limitations, potential biases, gaps & omissions

When certain data, especially regarding racial and socioeconomic demographics, is limited, omitted, or has gaps, when integrating multiple datasets, Group P may introduce a cascading bias into Wicked Problems Institute. As part of the ingestion and integration of data, WPI will develop a robust framework for the equitable treatment of data, including the mitigation of overfitting due to sampling bias, in order to eliminate systematic error of the data ingestion and integration pipeline. Furthermore, where it is not possible to outright eliminate bias or error, Group P will develop methods for the transparent explanation of the results, including an emphasis on digital equity.

Legal and ethical permissions for the reuse of data

Datasets will be primarily sourced from government agencies, which shall be reused within the bounds of respective licenses. This will also hold for any NGO-provided or proprietary data that is acquired.

Personally identifiable information

When combining datasets – especially when it comes to overlapping governmental datasets – it may be possible to de-anonymize data. Group P will ensure that all steps are taken to prevent this or make it as difficult as possible to target individuals using WPI.

Negatively impacted demographics

As noted above, certain demographics, and indeed certain regions demarcated by socioeconomic factors, may have limited coverage from data publishers. By failing to account for these omissions, WPI can exacerbate the problem; Group P will take all possible steps to ensure equitable treatment of every demographic.

Path to success through connectivity

Connectivity among all datasets and existing graphs could be realized in multiple ways, including linkages on the basis of specific existing physical features, or through more abstract notions like space, time and resolution. These connections should be considered on a case-by-case basis as they will differ depending on the types of data/graphs that are being integrated.

Group Q: Public Risk Analysis & Alert System
Powered by a Financial OKN

Group Q contributors focused on a business and finance OKN, built on publicly available data sources, to provide warnings and indicators of risks and opportunities for various industries and government agencies.

Contributors

Group Q contributors come from academia, government agencies, and the private sector. Each brought their unique experience and perspective. Contributors to Use Case B and their organizations are listed below.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Sam Chance	Cambridge Semantics
Vinay Chaudhri	JPMorgan Chase
Victor Zitian Chen	Fidelity Investments
Ying Ding	University of Texas at Austin
Oktie Hassanzadeh	IBM Research
Jay Pujara	University of Southern California
Louiqa Raschid	University of Maryland
Emanuel Sallinger	Vienna University of Technology and Oxford University
Ellie Young	Common Action

Champion

Oktie Hassanzadeh from IBM Research, was champion for Use Case Q. As champion, Dr. Hassanzadeh organized use-case group meetings and briefed contributors on group updates. Emanuel Sallinger, of TU Wien and Oxford University, and Louiqa Raschid, of the University of Maryland were Use Case Q Co-Leads.

Financial OKN for public good

A public service and/or tool that provides indicators and indices around risk factors (negative) and opportunities (positive) for industries / sectors. The service is powered by a business and finance OKN curated from publicly available sources (e.g. SEC filings, analyst reports), news reports, Web text, etc. A technical challenge is turning unstructured and structured data into useful knowledge; linking across resources and keeping the knowledge network updated.

Group Q give a number of examples of how such an OKN can serve the public good:

- **Baby formula supply chain issue.** This requires both information on the supply chains, as well as, crucially, financial information about the involved companies. Stakeholders include the DHS, the DOT, and many others. KGs/OKNs provide a natural way to integrate and reason over such an issue as the recent baby formula supply chain issue. In this case, Group Q sees an example of concentrate manufacturing. Similar examples can be seen in other industries, such as chip manufacturing.

- **Takeovers of strategic companies**. Crises such as pandemics or other stress situations make national companies of strategic importance vulnerable to takeovers, such as in the health, energy, defense or other sector. Preventing this requires understanding of both the legal and financial frameworks, as well as substantial reasoning power. Financial KGs/OKNs are a natural tool to provide such solutions. (example from the Central Bank of Italy: Bellomarini et al: Reasoning on Company Takeovers during the COVID-19 Crisis with Knowledge Graphs. 2020.) Stakeholders include government agencies and decision makers at all levels (SEC, DOC, DOJ, etc.)

- **PPP / U.S. Cares effectiveness**. Governments provide financial assistance during a crisis, ranging from loans to other financial guarantees. It is critical that governments also have the tools to judge the effectiveness of these investments (saving jobs; preventing bankruptcy). Since most such disasters are unexpected there are typically no a priori datasets that can be used for such analysis. KGs or OKNs about companies provide the needed infrastructure to create datasets on-the-fly to determine effectiveness of such financial support. U.S. Treasury, FRB, SBA are potential stakeholders.

- **Banking and company supervision**. Supervision of banks and other financial institutions is clearly an important task, as not only the great financial crisis has shown. Knowledge graphs and OKN capture the complex financial knowledge and data involved in this setting, and allow effective reasoning over them (example: Bellomarini et al.) Key stakeholders include all supervisory authorities and agencies, including those on banks and, in general, companies.

- **Climate finance**. In October 2021, the U.S. Financial Stability Oversight Council published a report describing climate-related financial systemic risks and outlining future work needed to analyze these risks. This crucially requires information on environmental models and financial models. Key stakeholders include the SEC.

Resources

Use Case Q identifies existing resources that could be incorporated into the financial OKN. These include relevant datasets, tools & infrastructure, and other expertise.

Datasets

Use Case Q focuses on current publicly available datasets for expediency. A successful financial OKN would combine both public and private / proprietary datasets.

Publicly available sources of data and knowledge for curating our business and finance OKN:

- U.S. SEC filings
- Publicly available analyst reports, for example:
  - NATO's Strategic Foresight Analysis reports
  - Reports from International Central Banks Resources with respect to international central banks (Central Bank of Italy/Oxford/TU Wien)
  - Knowledge graphs related to company control, takeovers, etc.
- Wikipedia & Wikidata as general-domain sources of knowledge.
- The Financial Industry Business Ontology (FIBO)
- Global Legal Entity Identifier Foundation
- Open Corporates

Tools & infrastructure

- Tools for knowledge extraction from text documents. Relevant tasks include: Entity Linking, Link Prediction, Table Extraction, Semantic Table Understanding, Causal Knowledge Extraction
- GPUs! Ideally, shared development environment (e.g. Python notebooks on the cloud) capable of handling the above tasks

Other expertise

A critical need for us is domain expertise. Ideally, Group Q needs experts from the financial regulatory agencies and financial analysts from banks and other investment groups.

- SEC (ongoing), Federal Reserve
- U.S. Small Business Administration (SBA)
- FinCEN in the U.S. Treasury

End-users and stakeholders

Government decision makers, analysts, and regulators can use this OKN to understand trends, monitor events, make investments to avoid crises, etc. This includes the financial health of businesses as well as the health of the labor market. Individual citizens and entrepreneurs may also use this OKN to make more informed decisions.

The financial OKN:

- Supports non-commercial or non-investment needs. Examples include the pandemic and accompanying supply chain challenges.
- Differentiates from tools that primarily support investors. Can we first identify the gaps in the tools developed for investors? How can we complement tools for investors? An example may be where there are sectors / trends that are not an investment opportunity.

The University of Maryland has a Center - **Small Business Anti-Displacement Network**. They are not tech-savvy or data-rich but they are well connected to a network of groups across the U.S. that provides services and support to small businesses. The planned financial OKN could be tailored to support the needs of such organizations. It can help to strengthen entrepreneurship and have a strong equity and diversity impact.

Figure 12 depicts an interconnected network of end-users and stakeholders as envisioned by Group Q contributors. Details on individual organization are given next:

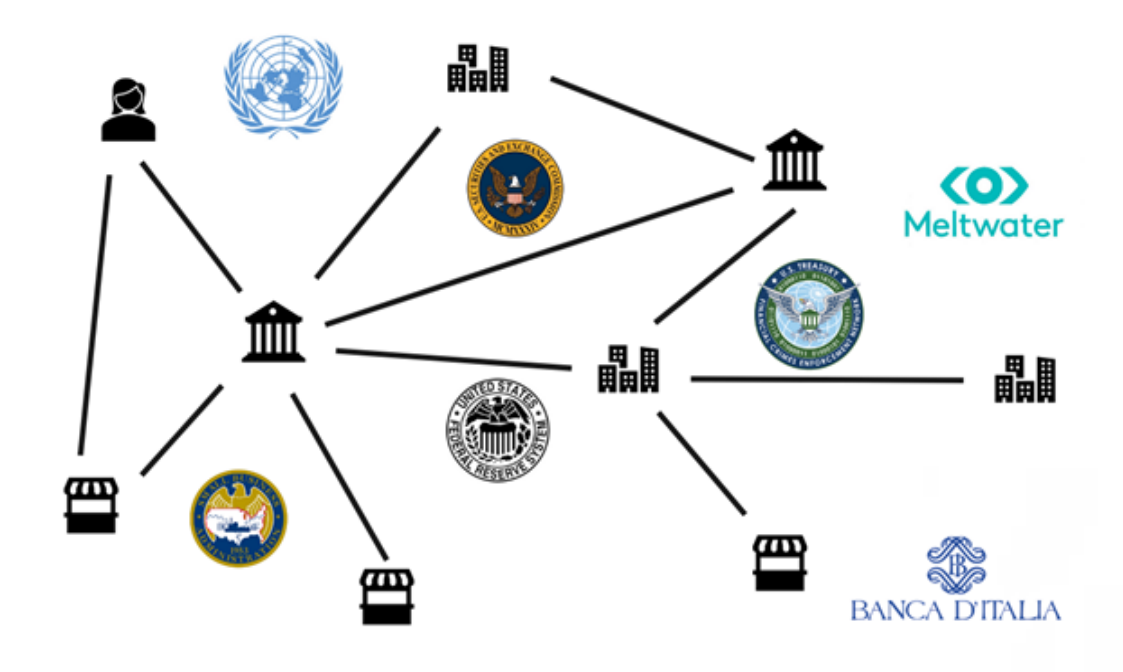


Figure 12: Network of end-users and stakeholders of the financial OKN

Federal agencies:

- **SEC and SBA**
  - Connection with Group K's interest in SEC's sustainability interest.
  - Interest in studying firms that will soon become public.
  - Detect fraud and identify outliers,
  - Use of standardized identifiers across agencies and sources
  - Incorporate news/social media into financial analysis using time-series of econometric data.

- **U.S.PTO**
  - Entity resolution — find all patents/trademarks from an entity.
  - Dynamic updates — company name changes (FB->Meta), address changes, and ownership changes.
  - OKN could provide tailored information, GUI, access, outreach/training, support to under-represented clients, e.g., small businesses that cannot purchase proprietary datasets or services.
- **Federal Reserve**
  - Identify the health of financial institutions.
  - Capture the macro-economic state of the U.S. economy.
  - Help re-use tools across the organization and leverage big data.
- **Government Accountability Office**

International agencies:

- **Central Bank of Italy**
- Key innovator in the European Financial Knowledge Graph space
  - Substantial investment in Knowledge Graphs for
    - Supervision
    - Anti-Money Laundering
    - Takeovers of strategic companies
    - Company control
    - Evaluation of financial and economic risks
    - Compliance
    - Substantial interest and track record in both applications and IT R&D
- **United Nations**
  - Strong interest in financial crime detection
  - Both in terms of Strategic Technology Solutions organization and in terms of Technology Innovation Labs
  - Private and public companies:
- **Media intelligence**
  - Example: Meltwater
    - Large interest in company knowledge graphs.
    - Aggregating news and other signals for financial entities such as companies.
    - Good example of financial knowledge graphs that include information on small or non-public companies.

Human-centered design

As part of the human-centered design concept, Group Q identified the following organizations as primary stakeholders:

- SEC (as key data provider)
- SBA (as U.S. small businesses would be one of the major beneficiaries)
- Central Bank of Italy

Ethics & social implications

The primary limitation of Use Case Q data sources is lack of high-quality sources of knowledge for the private sector and small businesses. Use Case Q contributors need to ensure that the data sources that are based on publicly available sources and mostly around public companies, are not biased in ways that would only benefit large corporations. As an example, a large company or a well-financed investment business would quickly find ways to ingest and exploit a financial OKN. A small business or a community organization fighting gentrification may find it difficult to utilize and exploit the financial OKN. One way to address this challenge is to make sure that there are APIs and UIs to make the financial OKN more accessible. A simple example may be to provide downloaded data in CSV files instead of JSON.

Group Q contributors should also make it very clear that business and finance OKN powered service relies on such sources and such biases may exist. Use Case Q contributors should allow mechanisms for feedback to data providers (e.g., U.S. SEC) in case biases and potential negative impacts arising from the contents provided by them are identified.

Furthermore, it is a desirable goal that the Knowledge Graph / OKN itself contains at least the knowledge on applicable laws and regulations, especially with regard to privacy (PII, etc.), making them operationalizable as part of the OKN. Some examples include Bellomarini et al., Financial Data Exchange with Statistical Confidentiality. EDBT 2021. As far as possible, further ethical and legal standards should be included and made operationalizable as well.

Path to success

Group Q contributors identified connectivity as essential to the ongoing success of the Financial OKN. There are clear and direct connections to other uses cases, in fact many of the key driving examples require these connections:

- **Supply Chains** (Group A): Supply chains typically had a focus on logistics and products. The underlying ownership and financial supply chains are also critical elements to obtain a comprehensive overview of the supply chain.
- **Decision Support for Government Leaders** The financial OKN can support financial regulators, e.g., protecting companies that are of strategic national importance, or helping governments make decisions at a local level, e.g., what communities and businesses can benefit from financial assistance during a pandemic.

- **Education** (Group N): Financial KG/OKN cases provide very clear and approachable examples for education, and have been used in this area frequently. As collaboration with Group N, we conducted a survey of both education and financial KG/OKN stakeholders, emphasizing the synergies.

Group R: Bias in Research Priorities Knowledge Monitor

Group R contributors propose to build a knowledge graph that will aggregate micro-level data on research activity in the life sciences with particular attention to contemporary and historical sources of bias and gaps.

Contributors

Group R contributors are listed below, along with their organization.

CONTRIBUTOR'S NAME	CONTRIBUTOR'S ORGANIZATION
Luis Amaral	Northwestern University
Ying Ding	The University of Texas at Austin
Jennifer Dunn	Northwestern University
Jianxi Gao	Rensselaer Polytechnic Institute
Matthew Lange	IC Foods
Adam Pah	Northwestern University
Douglas Rao	National Oceanic and Atmospheric Administration
Ilya Zaslavsky	University of California San Diego

Champion

Luis Amaral, Erastus Otis Haven Professor of Chemical and Biological Engineering, Northwestern University, was champion for Use Case R. As champion, Prof. Amaral organized use-case group meetings and briefed contributors on group updates.

Identifying sources of bias and gaps in life sciences research

In many domains of research in the life sciences, broadly construed, we know that there are disparities in how research is funded, how researchers are funded, and how topics rise to disciplinary importance and prestige. This introduces different types of bias into the field more broadly, including creating

barriers in what types of knowledge are elevated and pursued. This use case aims to systematically aggregate data about how science is being conducted and funded to identify and alleviate these biases. For example, research has demonstrated that National Institutes of Health (NIH) funding for specific diseases is not aligned with the burden of the disease. Burden of disease is defined as the impact of a health problem, as measured by prevalence, incidence, mortality, morbidity, extent of disability, financial cost, or other indicators.

Figure 13 shows differences between NIH predicted funding and actual funding for two fiscal years (1996, light blue, and 2006, dark blue). Predicted funding is based on the burden of disease and the actual NIH expenditures for two fiscal years. Negative values indicate lower funding than predicted. The x-scale axis is in millions of dollars. Remarkably, there are three conditions — COPD, and injuries for which the funding deficit exceeds $0.5 billion dollars.

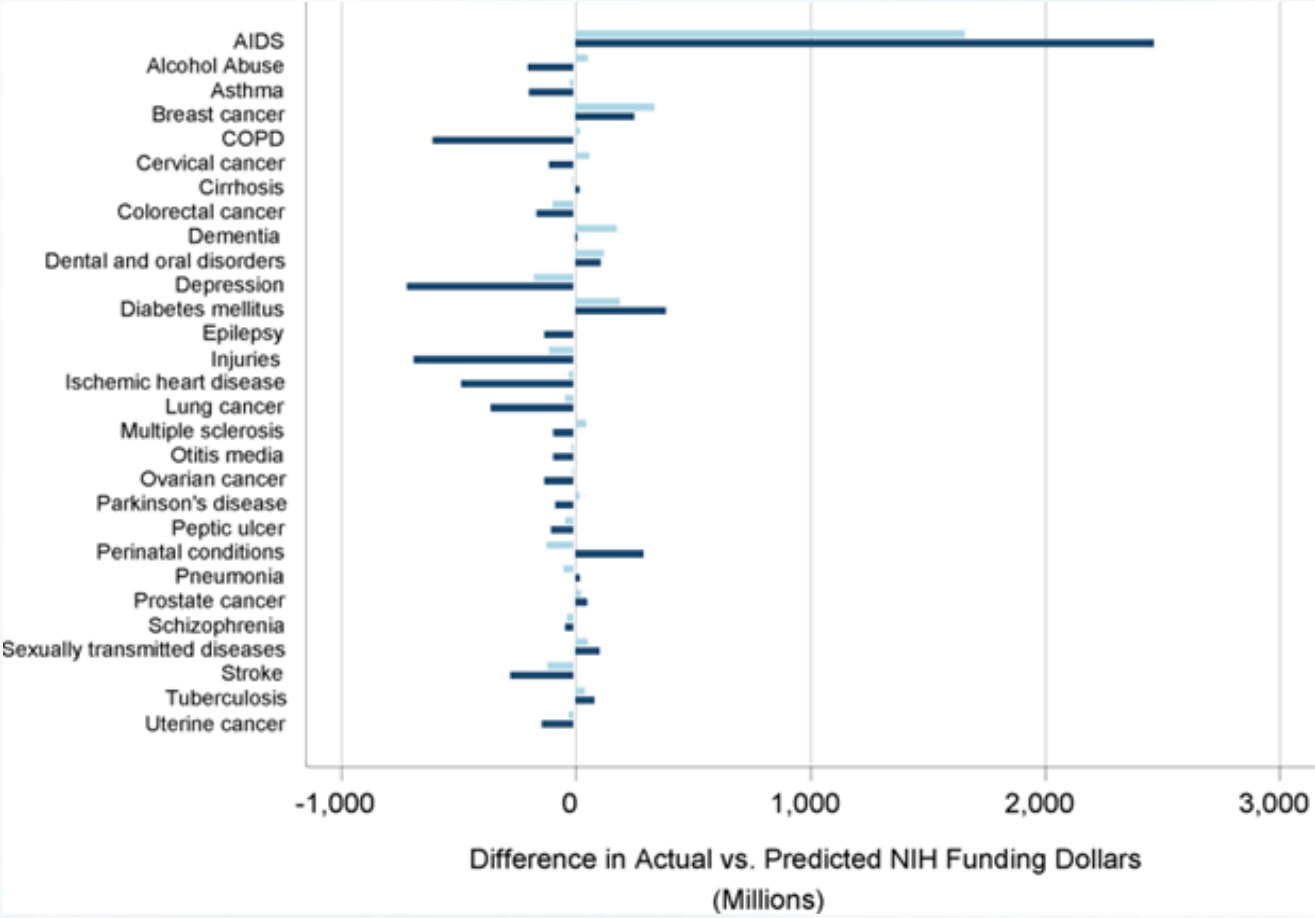


Figure 13: Difference between NIH predicted funding based on burden of disease and the actual NIH expenditures for two fiscal years (1996, light blue, and 2006, dark blue)

Misaligned funding is not, however, the only issue where the national research enterprise is far from optimal. Research has demonstrated that non-scientific reasons delay implementation or adoption of important research directions. For example, despite the importance of diet to health span, it was only in 2015 that the NIH established an Office of Nutrition Research, and only in 2020 that a Nutrition Strategic Plan was developed. One may ask what was the role of structural sexism – dietetics is a female-dominated profession – in such a delay? Or what was the impact of the absence of commercialization outcomes for dietary recommendations? Merging these data resources will enable the systematic study of how bias perpetuates and affects funding directions and decisions throughout the entire funding and research life cycle.

While these types of issues are not restricted to biomedical research, the life sciences offer a case study where the implications are immediately apparent. They are also a case where there is already ample data, making it easier to develop a minimum viable product.

For these reasons, the proposed OKN will aggregate micro-level data on research activity in the life sciences with particular attention to contemporary and historical sources of bias/gaps.

Goals & objectives

Use Case R goals and objectives are to

- Identify overlooked research approaches,
- Enable government transparency
- Facilitate a more equitable and robust scientific funding allocation while accounting for historical, disciplinary, technological, and economic biases of scientific funders, and
- Enable individual researchers to find open research niches.

Resources

Development of ontologies characterizing scientific modes of inquiry, their processes and technologies, in addition to the diversity, access, and motivations of investigators using these modes of inquiry, as well as the populations who benefit from the research; will be necessary in order to capture specific classes of bias, and instantiate them such that tradeoffs can between alternate courses of research can be compared.

The life sciences are flooded with data sets, ontologies, and approaches for extracting information from the literature (**see figure 14**). Many of these resources are actively maintained and curated.

Figure 14: Data sets, ontologies, and tools for extracting information from life sciences literature.

Burden of disease: IHME at University of Washington collects and processes data
Research expenditures: The National Center for Science and Engineering Statistics collects and processes national data at the institution level. The NIH collects and processes national data at the institution and PI level.

Research publications: PubMed comprises more than 33 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites. These data have been connected to the Microsoft Academic Graph and updated to 2021.

Ontologies

- pubTator provides information on mention of biomedical entities in more than 30 million biomedical publications.
- BioBERT is a pre-trained biomedical language representation model for biomedical text mining
- MeSH (Medical Subject Headings) are standardized keywords. The majority of articles in PubMed (over 90%) have been assigned MeSH terms to provide information on the content of the articles. MeSH terms are manually assigned by indexers of the National Library of Medicine.
- The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes.
- Disease ontology (https://disease-ontology.org/),
- CheMBL and drugbank for drugs.
- EXPO is an ontology of scientific experiments that links the Suggested Upper Merged Ontology with subject-specific ontologies of experiments by formalizing the generic concepts of experimental design, methodology and results representation.

Data needing to be developed:

While there are a multitude of resources, not all needed data is available. The area where this is most

pressing is in developing a way to extract scientific approaches and disciplines used at different levels. Specifically, we need to

- Identify biomedical researchers' expertise in terms of
  - approach (computational, experimental, statistical, theoretical, etc.)
  - discipline (biology, CS, chemistry, physics, psychology, etc.)
  - experience (time since publication of first paper)
- Identify expertise brought to bear upon production of:
  - scientific papers
  - projects/grants
  - public data or software resources

Other expertise

The study of science of science, science policy, and innovation is a burgeoning new field that brings together informaticians, epistemologists, sociologists, psychologists, historians of science, economists, computer scientists, and data scientists. This new OKN would enable transformative studies of the national research enterprise and provide vital information to different stakeholders (see **figure 15**).

Figure 15: Potential stakeholders for use case R OKN

End-users & stakeholders

The platform developed by Use Case R group would be used by all major stakeholders of the biomedical research and development enterprise. For example:

- Scientists could be better able to identify areas of great health importance that lack appropriate attention
- Scientific societies would be able to match expertise to overlooked areas of application of that expertise
- Funding agencies could better track alignment between research priorities and spending
- Decision makers in the executive and legislative branches at the federal and state level would better be able to evaluate between the needs of the nation and on what funds are spent.

Stakeholders

The National Center for Science and Engineering Statistics, the National Center for Biotechnology Information at the National Library of Medicine are key partners for the development of the OKN. They already collect, analyze and distribute much of the data that would be ingested by the OKN and already have processes in place for the storage and distribution of data of national importance.

Human-centered design

To understand what users and stakeholders need, we consulted with existing stewards of data related to scientific funding decisions and scholarly experts in the science of science. In these discussions we identified that the missing gap is a resource that actually unites the various, existing biomedical databases already identified. The data extraction in certain databases, normalization of terms, and mapping entities between datasets is the primary roadblock to building a systematic understanding of how funding and research activity is related to the current burden of disease.

Ethics & social implications

All primary data to be used in developing the MVP is already in the public record. However, it is not integrated as a knowledge network which limits its utility in guiding policy-makers and funding institutions in defining their priorities and researchers in identifying new opportunities.

To fully realize the vision of the bias in research monitor it will be necessary to develop and use AI to extract data from the primary research sources. This is a clear step where bias can be introduced based on the performance of the tools. To combat this, these models will have their performance evaluated in an objective, independent manner with public datasets that are created, maintained, and available to the community along with the models.

The OKN itself will be updated through discrete releases (e.g., annual). This will effectively version the contents of the OKN itself that are tied to both increased data availability and model improvements. The prior OKN content versions will also remain available indefinitely in order to support the broader needs of open science and reproducibility in research.

Path to success

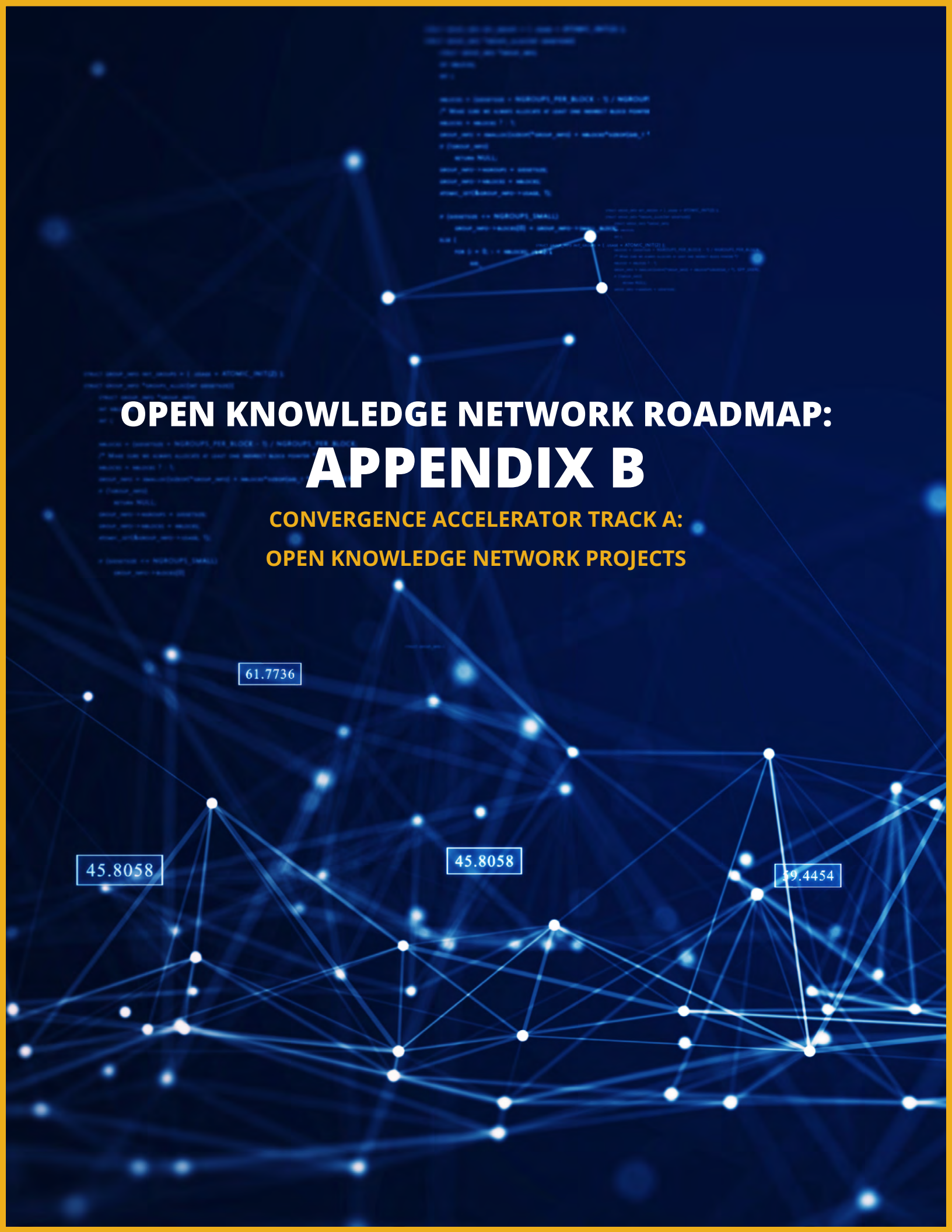
Group R contributors identified connectivity and scalability as the essential factor in achieving success.

Connectivity

In the long term, the MVP could be integrated into existing efforts that already collect and organize these types of data, such as the National Library of Medicine or the National Center for Science and Engineering Statistics. The initial difficulty is primarily in building the initial software tools and infrastructure that form the basis of the MVP OKN.

Scalability

There is also the need for scaling the OKN to other additional areas, such as: Materials Science and Engineering, Social Behavior, and Sustainability and Resilience research. However, expansion to these areas will require more additional work since tools and data resources are not nearly as developed as for biomedical research. These topical expansion will require greater support and investment from subject matter experts in those areas.



OPEN KNOWLEDGE NETWORK ROADMAP:

APPENDIX B

CONVERGENCE ACCELERATOR TRACK A:

OPEN KNOWLEDGE NETWORK PROJECTS

TABLE OF CONTENTS

CONVERGENCE ACCELERATOR TRACK A OKN PROJECTS

Appendix B provides an overview of five Open Knowledge Network (OKN) projects supported by the NSF **Convergence Accelerator Track A Program**. The **Convergence Accelerator** Program builds upon basic research and discovery to accelerate solutions that have substantial societal impact. The program funds teams to address societal challenges through convergence research and innovation. This convergence approach, including a strong focus on human-centered design, brings together the disciplines, approaches, and organizations that are needed to create innovative solutions, refine prototypes, and then implement useful tools and deliverables.

The goal of Track A - Open Knowledge Networks was to enable new modes of data-driven discovery that help advance the progression from data to knowledge. The 21 **Phase 1** projects funded under Track A identified challenges in different topical domains (e.g. flooding, personalized medicine) as well as "horizontal" challenges that cross multiple domains (e.g. incorporating geospatial data, managing provenance). Phase 1 grantees refined their ideas by engaging with stakeholders using human-centered design methods and expanded their multidisciplinary and multi-institutional teams to create effective deliverables.

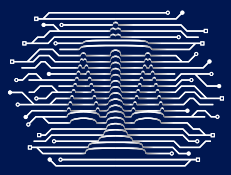
Subsequently, 5 of these 21 projects were selected for a two-year **Phase 2** effort. While user needs were explored in Phase 1, a Minimal Viable Product (MVP) and of a clear set of product deliverables was to be developed in Phase 2. The five Open Knowledge Networks phase 2 efforts include:

- **Systematic Content Analysis of Litigation EventS (SCALES) - OKN**: Led by Northwestern University, the SCALES open knowledge network is designed to be a public resource to help provide insights based on judicial court records. SCALES is creating tools to decode court records and transform this data into actionable information that aids a variety of uses, including legal scholars, journalists, policymakers, judiciary, and citizens.
- **Scalable Precision medicine Open Knowledge Engine (SPOKE)**: Led by University of California, San Francisco, this Biomedical Open Knowledge Network connects millions of biomedical facts including molecules, pharmacological compounds, organs and diseases, food nutrients, and more. Centered around knowledge representation and reasoning, the team is developing applications using graph theory, advanced visualizations, and real-world clinical evidence to advance drug development and precision medicine.
- **Urban Flooding Open Knowledge Network (UF OKN)**: Led by the University of Cincinnati, the UFOKN is addressing urban flooding impacts to assist decision-makers and urban planners in real-time response and long-term planning.

- **KnowWhereGraph (KWG)**: Led by the University of California, Santa Barbara, KWG provides knowledge graph and geo-enrichment services for environmental intelligence applications. The solution enriches data with pre-integrated custom-tailored knowledge about any locale of interest, thereby reducing the time to find, combine, and reuse data. The initial application areas are focused on decision support related to food systems, supply chains, and humanitarian aid, but can easily be expanded to other application areas as well.
- **Knowledge Network Programming Infrastructure (KNPI)**: Led by the University of Michigan/ MIT, KNPI is building infrastructure for constructing novel OKNs and OKN-powered applications. This solution provides tools to make the creation and maintenance of high-quality datasets and apps more cost-effective and more widely accessible.

The many accomplishments of the five Track A - OKN teams that advanced to Phase 2 are described below, including the integration efforts they have underway to help create elements of a single, nonproprietary, shared knowledge infrastructure. You can find information on these efforts and the other topical tracks supported by the Convergence Accelerator in the **Portfolio**.

SCALES - OKN



The SCALES Open Knowledge Network is an AI-powered platform that provides users access to judicial opinions and insights. It has achieved its project goal of making these transparent to anyone seeking this information.

Project goals and accomplishments

The goal of the SCALES OKN is to bring transparency to the systems and processes of the U.S. courts. Transparency will help ensure that systems and processes are fair, efficient, and accurate. The 94 federal courts have a profound effect on the administration of law in the United States. They decide cases that range from civil rights issues to nationwide product liability cases that affect every single American.

However, the reality is that almost everything we know about how the federal judiciary works comes from the written opinions of judges. These opinions are issued in less than 10 percent of cases — creating a vast dark matter of litigation that academics, policymakers, and the public know little about. To accomplish the goal of transparency, we have built an AI-powered data platform. The platform makes the details of the federal judiciary and insights into how it works available and accessible to everyone.

Through the creation of the SCALES OKN we have developed software and analytics that reveal how the system currently works. This transparency has helped legal scholars, legal aid organizations, and the courts themselves better understand the system and how it could be improved. We developed an analysis on how variable court decisions are. The analysis, included in a report targeting courts and lawyers' associations, shows that the variability of court decisions often has negative effects.

Specifically, the decisions on plaintiff requests to waive the filing fees burden limited-income plaintiffs who cannot pay them. A case in the district courts and the Chicago Council of Lawyers and Appleseed Network have used this report to advocate for a policy change in how these requests are treated. We have also prepared and delivered this report directly to two other federal district courts.

Similarly, we have examined the frequency of motions to seal in intellectual property cases. We also looked at how this frequency varies across districts. The New York University Technology, Law, and Policy Clinic and the Electronic Frontier Foundation included our findings in their brief regarding sealing practices in the courts.

The technology that we have developed to ingest court data now supports a number of other projects. The Civil Rights Clearinghouse is a searchable resource for information and documents relating to civil rights litigation and now uses SCALES software to automatically ingest court records and documents. Similarly, the Full Disclosure Project at the National Association of Criminal Defense Lawyers relies on the SCALES technology stack to automatically download federal court records and populate the project database. The Full Disclosure Project is built to help criminal defense lawyers have a searchable and connected database of police misconduct that aids them in advocating for their clients in court.

Importantly, the goal of the SCALES OKN is to build the connective tissue from court records to other data resources. This will enrich the OKN and advance what users can glean from the data. Towards that end, we have integrated the Federal Judicial Center's open data on judge biographies and metadata for filed cases. We've also created the linkages between the corporate parties involved in cases with the Security and Exchange Commission's EDGAR database. These connections allow users to ask and answer intricate questions about who is involved in a case and how these cases reach a resolution.

Videos, websites, and other resources

The following resources provide additional information about SCALES. It also provides a link for users to register for SCALES to access and explore its information.

- Website: **SCALES OKN**

- **Registration for SCALES OKN access**

- **Science: How to build a more open justice system**
  A policy forum article published in Science by the SCALES team. It describes why systematic analysis of court records is needed. It also describes how it can be accomplished in a manner that benefits the judiciary and the public.

- **ACM Digital Library. From data to information: automating data science to explore the U.S. court system**
  An article that describes the development of the SCALES OKN application prototype and its function.

- **IEEE Abstract. PRESIDE: A Judge Entity Recognition and Disambiguation Model for US District Court Records**
  An article that describes the SCALES OKN models to identify judges by name. Models use free-text. They disambiguate judges' names into a single true entity (i.e., so that all mentions of Judge Jane Doe are connected across all cases).

- **Wiley Online Library. AI Magazine: The Promise of AI in an Open Justice System**
  An article that describes what the SCALES OKN is. It details how AI-driven tools will grow in utility in the legal data domain and help improve the justice system overall.

- **Tear Down This Judicial Paywall - WSJ**
  Opinion from SCALES team members in the *Wall Street Journal.* They advocate for free access to federal court data.

- **Opinion: Texans shouldn't have to pay for court records**
  Opinion from Amy Sanders in the *Houston Chronicle*. Texans should have free access to court records to assess local issues related to social justice.

- **Scales-dash**: *Analysis of Sealing Activity in Patent Cases*
  A public interactive report on the variation frequency of court documents sealing across courts in intellectual property cases.

- **Scales-dash**: *Crosswalking Pacer to the IDB*
  A public interactive report that documents the issues in the official Federal Judicial Center Integrated DataBase when the data column is concerned with issues that affect indigent parties.

- **Civil Rights Litigation Clearinghouse**
  The Civil Rights Litigation Clearinghouse has integrated the SCALES software to automatically download data from PACER into their platform.

- **GitHub - scales-okn/PACER-tools**
  A public repository of our open source software to obtain and parse PACER data. This software has been used directly by multiple federal public defender's offices and researchers.

- **Materials related to SCALES OKN research pieces and blog posts**
  A public repository of all code and data used to produce research publications.

Track integration goals and accomplishments

Our goal is to integrate the work done by KNPS on data-provenance tracking with the Saytrn system developed at SCALES. Satyrn is an information system designed to answer questions about a knowledge network. It does this by mapping information requests from the user into a data analysis and results presentation. Importantly, the system also analytically infers what the proper form of the answer is (whether that be textual or a graph) and delivers the corresponding result. The focus of the work is to provide users with frictionless access to both the information that can be:

- Derived from a data set, and
- Qualified based on the meta-data about its provenance.

The concrete software engineering goal is to integrate the SCALES Satyrn analytics environment with provenance data enabled by the KNPS system. This integration effort would make systems more useful and the resulting software work will comprise two components:

1. Automate Satyrn configuration by exploiting past KNPS provenance recordings, thereby saving human work for each novel data analysis.
2. Provide easy log-driven KNPS provenance publishing hooks in Satyrn so downstream users can reliably review past data analyses.

This software work will be valuable on its own merits, while also furthering higher-level goals:

- Demonstrating implicit work sharing:
  - Among provenance users, and
  - Low-overhead provenance recording by app developers.

The initial work on this integration has focused on defining the structure of the meta-data that will be shared between these two systems and the identification of integration entry points. Our starting point in this work is to use the KNPS metadata as an additional source in the configuration of Satryn. Our longer-term goal is to integrate the configuration layer into the KNPS system itself. The integration will include hooks.

This work is crucial for the broader Convergence Accelerator cohort. The fully realized integration of the Satyrn and KNPS systems will support the effective ingestion, integration, and maintenance of a wide range of data sets. Users, even those without data or analytics skills, can directly access the information that is contained within those data and their analysis. The resulting platform could open the door to a wide range of data sets. Users who would otherwise have to rely on external technical skills or staff could access data sets for more effective utilization.

We believe that this is an opportunity to explore and shape the future of data infrastructure. Data are increasingly utilized in many facets of daily life, and yet the processes for data ingestion, cleaning, reconciliation, and analysis are often bespoke and opaque. As an example, we recall the information challenges of the early stages of the COVID-19 pandemic. Public health officials needed to know the location and status of infections, medical infrastructure, and healthcare workers. While the necessary data existed, mechanisms for aggregating, normalizing, and explaining them did not. We believe this project can help to establish the standards for tools and workflows that would obviate such problems.

SPOKE

SPOKE — the Scalable Precision medicine Open Knowledge Engine — is a biomedical knowledge network. It was created by integrating several specialized databases. It is to be further developed into a multi-domain network of biomedical knowledge and data on a massive scale.

Project goals and accomplishments

Human physiology and pathology are governed by molecular pathways of daunting complexity. These processes are represented in disparate scientific datasets siloed across thousands of public repositories. Siloed datasets make it nearly impossible for researchers to:

- Utilize the entire body of data and factual knowledge, and
- Connect the dots across the domains of specialization in biomedicine.

Yet, Big Data must be converted into Big Knowledge if we are to harness the data revolution. The SPOKE team is working to develop a multi-domain network of both biomedical knowledge and data at a massive scale. These will be validated with real clinical evidence that will enable:

- Investigation of connected biomedical concepts, and
- The emergence of new knowledge

This will be accomplished by the convergence of seemingly disparate knowledge repositories and data sources.

As part of the NSF Convergence Accelerator program, SPOKE integrated several specialized databases. Integrating these databases resulted in a biomedical knowledge network. That network represents 23 million concepts with 50 million connecting relationships. The network is accessible both programmatically (via API) and through a graphical user interface. This allows investigators to reason across a vast, dispersed body of scientifically accepted pathways governing human health. The multidisciplinary project team is collaborating on:

- Network architecture
- Data modeling, and
- Scaled visualizations

The team is also collaborating on applications of the network. These include research questions in the areas of drug discovery and disease diagnosis and management.

The team has demonstrated the utility of the network and the validity of a network-based analysis approach. This approach enables the network to generate hypotheses and provide prognostic information about patients. For example, in March of 2020 SPOKE integrated the newly published SARS-

CoV-2 Interactome into the network to explore some potential pathways of viral activity. They were able to identify pathways and predict promising therapeutic measures. These were later confirmed experimentally in the literature.

The team also launched a company, Mate Bioservices, to commercialize development of a suite of products powered by SPOKE. This suite of products will enable wide dissemination. It will maximize the utilization of this biomedical OKN for the benefit of society as a whole.

By the end of the project period, we will release a web-accessible network visualization tool aimed at citizen scientists. We will produce a report outlining recommendations for mitigating risk associated with ethical, legal, and social implications of network use. And we will deploy the first clinical decision support system (alpha) in the UCSF neurology practice with real patients.

Videos, websites, and other resources

The following resources provide additional information about SPOKE.

- Website: **SPOKE Informational**
- SPOKE **Neighborhood Explorer**
- **Mate Bioservices**
- **A5: Biomedical Open Knowledge Network // Project Video**
- YouTube: **Biomedical researcher**
- YouTube: **Pharmaceutical developer**
- YouTube: **Clinical service provider**

Track integration goals and accomplishments

SPOKE and KnowWhereGraph (KWG) are working on a functional integration plan that will enable users to navigate from one KG to the other when knowledge on that specific area is supported by the partner graph.

While KWG has formal models (i.e., ontologies) for hazards and environmental data layers, including storms, floods, road networks, and health statistics of an area's population, it does not provide mechanistic information on how these could affect individual and population health. Interestingly, this

is exactly what the SPOKE graph contains, providing deep information and connections across diseases, symptoms, genes, drugs, and beyond.

This API-based integration will enable a full service and graph stack for events on the ground and the needs of the affected population. It will drastically improve the matching of experts. At the same time, this project will render both project graphs even more powerful. It will do this by integrating region-specific public health information with detailed data about diseases, their etiology, and recommended treatments. For instance, SPOKE would immediately gain access to making discoveries that involve regional, environmental, and social determinants of health while KWG could add in-depth medical knowledge to its use cases around humanitarian relief and food supply/safety.

Successful API-based integration between SPOKE and KWG will require three steps:

1. Alignment of vocabularies (e.g., taxonomies and ontologies, also called KG schema) used in both graphs.
2. Coreference resolution: Identification of in-common graph nodes and declare their global identifiers (URLs) as equivalent.
3. Establishing a workflow that enables actual queries to their public endpoints, as well as for the usage of these endpoints.

Access to contents across both graphs will provide users with a substantially deeper pool of facts to discover and query both vertically (by adding more data layers) as well as horizontally (by containing more data per layer). In addition, closely integrating the semantic representations (ontologies and design patterns) of the domains within our graphs will enhance interoperability, as well as tie abstract knowledge (such as the Disease Ontology) to concrete data about places in ways that are relevant and immediately impactful for users.

Urban Flooding Open Knowledge Network

The Urban Flooding Open Knowledge Network (UFOKN) is an open and shared infrastructure that enables optimal disaster mitigation and long-term resilience planning.

Project goals and accomplishments

Modern cities are a complex and interconnected system of engineered, natural and social systems. This interconnected system can be conceptualized as a network of networks, or a Multiplex. The multiplex includes the following:

- Power grid
- Transportation network
- Natural surface water and groundwater systems
- Sewerage and drinking water systems, and
- Inland navigation and dams

All are intertwined with the socioeconomic and public health sectors. While the Urban Multiplex connected, the data that have been collected for decades across multiple sectors have remained siloed. That is the main barrier to optimal disaster mitigation and long-term resilience planning. The Urban Flooding Open Knowledge Network (UFOKN) is breaking these data silos. It is an open and shared infrastructure that provides an information backbone for owners, operators, and consumers of various subsystems of the Urban Multiplex during flooding, and for long-term resilience planning.

UFOKN currently holds data on over 140 million critical assets across the continental US including buildings (residential, commercial, industrial, air/water/rail terminals), underground fuel storage tanks, superfunds and the power grid, with more data (e.g. roads and highways, communications infrastructure, wastewater outflow locations, socioeconomic data) being continuously added.

We now have reached the capability to generate real-time flood forecasts at these assets across the continental U.S. These achievements enable critical value adds such as a socioeconomic analysis of regional flood impacts and evacuation routing recommendations. To this end, we have developed a Computable General Equilibrium (CGE) model that assesses the broader economic impacts of flooding on economic activities. These include industry output, employment, and value-added and household income. We generate evacuation routing recommendations using real-time flood forecasts and an agent-based model.

The next phase of the UFOKN project will see a public launch of a series of data products aimed at a wide range of users — from emergency responders to urban planners, utility managers, local and state governments, researchers and the general public.

Videos, websites, and other resources

The following resources provide additional information about UFOKN.

- Website: **Urban Flooding Open Knowledge Network**
- YouTube channel: **UFOKN - YouTube**
- Twitter: **Urban Flooding Open Knowledge Network**
- Article: **Knowledge graphs to support real-time flood impact evaluation - Johnson - 2022 - AI Magazine - Wiley Online Library**

Track integration goals and accomplishments

The way hazards impact communities, infrastructure, and the natural environment depends to a large degree on their connectivity, as well as on their antecedent conditions. For example, the 2018 Camp Fire — the deadliest, most destructive and expensive disaster that year worldwide — was triggered by a small fault in the power grid. But another major contributor, which did not get as much attention, was the prolonged regional drought. The fire destroyed the infrastructure in the region, including buildings, utility networks, and healthcare systems. It caused cascading infrastructural and socio-economic disintegration across urban and rural communities. A large proportion of those displaced from the devastated rural community of Paradise moved to the nearby city of Chico. This increased the population by 20% virtually overnight, overwhelming all available resources.

In order to predict and mitigate such cascading events, we need information systems that integrate and harmonize data from disparate sources. They must also provide actionable information to decision-makers from local and regional to state levels. UFOKN and KWG are collaborating to develop such a system.

UFOKN holds high-resolution data (sub-kilometer, feature-level (e.g., exact location of buildings), while KWG's holdings are at a resolution that is fine-grained from an environmental intelligence perspective, but low for building-level impact modeling in UFOKN. We will develop bridges across geospatial data of varying resolution at a continental US scale. Second, we will develop a common-hazards ontology based on the current KWG version that supports a more fine-grained axiomatization of extreme events and their impacts. Collectively, this work will produce capacity to connect OKNs across various spatial and temporal resolutions and filter relevant connections and knowledge quickly and at scale. We will demonstrate this with a use case based on historic events during Hurricane Harvey in 2017, which we expect will open up further development of ontologies for a broader community.

KnowWhereGraph

The KnowWhereGraph is a cross-domain knowledge graph that supports data-driven analytics and decision-making pertaining to natural disasters and other threats to the environment, industry, and the financial sector.

Project goals and accomplishments

KnowWhereGraph aims to provide area briefings for any place on Earth within seconds. It will answer questions such as "What is here?", "What happened here before?", "Who knows more?", and "How does this region or event compare to other regions or past events?" As a cross-domain, FAIR principles-based knowledge graph for environmental intelligence applications, our current pilots include:

- Disaster relief
- Supply chain management
- Commodity trading
- Financial risk assessment
- Environmental, Social, and Corporate Governance

Our KnowWhereGraph supports data-driven analytics and decision-making by providing (1) a 12 billion facts-strong open knowledge graph that interlinks over 30 cross-domain data layers, (2) a pattern-based suite of expressive ontologies, and (3) a set of geo-enrichment services that enable rapid access to the graph from within the comfort of Geographic Information Systems such as ArcGIS and QGIS.

The team has made significant progress over the past two years in developing the KWG graph using expressive ontologies that connect: multi-source data in terms of disaster, air quality, climate hazards, crop history, soil characteristics, experts and expertise, administrative boundaries, health, transportation infrastructure, and so forth.

Overall, our graph provides 10 different kinds of geographic identifiers and over 20 data layers that provide millions of past and present facts about any of these regions, be it cities, lakes, or agricultural fields. We expect this number to continue growing as more automated graph generation and integration approaches are being developed.

Our team members are also pioneers in developing spatially explicit machine-learning models to provide GeoAI-ready data to empower intelligent decision-making. Combining classical deduction and constraints that checks representation learning enables us to serve a wide range of services to our partners. Services include similarity search, outlier detection, enrichment of existing data, alignment to other graphs, recommendations, link prediction and so on.

Recently, we centered our efforts on supporting project verticals, including the disaster relief subteam,

to assemble quickly needed datasets for rapid disaster response and evacuation after major devastating events, such as hurricanes, have occurred. We are also developing graph solutions for understanding and sustaining food supply-chain resilience.

Each of our pilots comes with bespoke user and query interfaces to accelerate the transition from data to knowledge. We also serve the entire graph openly in the form of a SPARQL query endpoint for developers as well as via our faceted search interface for easy exploration. As a technology-driven project, our goal is to demonstrate how novel geospatial solutions can inform downstream stakeholders from industry, nonprofits, and government agencies.

We are now deploying our graph and services to an increasing set of partners and looking for new opportunities to apply our methods to new use cases.

Videos, websites, and other resources

The following resources provide additional information about the KnowWhereGraph.

- Website: **KnowWhereGraph**

- **KWG - Pilots**
  Pilot programs utilizing KWG

- **KWG - Tools**

- **Know the Graph**
  Link to the graph, faceted search, and query endpoint

- **Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence - Janowicz - 2022 - AI Magazine - Wiley Online Library**
  Janowicz, K., P. Hitzler, W. Li, D. Rehberger, M. Schildhauer, R. Zhu, C. Shimizu, et al."Know, Know Where, KnowWhereGraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence." AI Magazine 43 (2022): 30–39. **https://doi.org/10.1002/aaai.12043**.

Track integration goals and accomplishments

KnowWhereGraph and SPOKE are working on a functional integration that will enable users to navigate from one knowledge graph to the other when knowledge on a specific area is supported by the partner graph. This effort has already been described above under the SPOKE project.

Access to contents across both graphs will provide users with a substantially deeper pool of facts to discover and query both vertically (by adding more data layers) as well as horizontally (by containing more data per layer). In addition, closely integrating the semantic representations (ontologies and design patterns) of the domains within our graphs will enhance interoperability, as well as tie abstract knowledge (such as the Disease Ontology) to concrete data about places in ways that are relevant and immediately impactful for users.

KnowWhereGraph and Urban Flooding OKN are collaborating to develop an information system that integrates and harmonizes data from disparate sources, and that provides actionable information to decision-makers from local and regional to state levels. This joint effort was already described above under the UFOKN project.

Knowledge Network Programming Infrastructure (KNPI)

OKN Infrastructure

Knowledge Networks are a novel and potentially transformative form of data, but building applications on top of them is too difficult, time-consuming, and expensive. We are building a Knowledge Network Programming Infrastructure that makes it far easier to build novel knowledge-powered applications, while also improving the knowledge resources themselves.

Project goals and accomplishments

Open Knowledge Networks are compelling artifacts that enable a range of novel applications, many of which are embodied in the ideas in the Innovation Sprint. Despite OKNs' recent growth, building a novel knowledge network requires substantial engineering effort. Not as many people and organizations benefit from OKNs as should be possible. The core objective of the Knowledge Network Programming Infrastructure project is to make Knowledge Networks easier to build and debug. We do that in several ways:

- Extraction infrastructure software, which makes it easier to populate an OKN with the large set of facts that can be derived from documents
- Provenance infrastructure software, which makes it easier to debug OKNs by creating a queryable record of data operations that is low-cost and works across institutions
- Program synthesis tools that make it easier to induce new OKNs

KNPI extraction software

This software focuses on obtaining OKN facts from natural language documents, in particular scientific documents. Scientific information extraction has been an area of research and engineering activity for at least several decades, but has historically required enormous effort to obtain high-precision and high-recall results. Our extraction software aims to yield high-quality results while dramatically reducing the amount of human effort required. The KNPI extraction suite has four important components.

1. The **VIsualLayout, or VILA**, system is a trained model that breaks documents into visually coherent pieces that are semantically self-contained; for example, article text and image captions will be grouped separately, even if they are geometrically quite close to each other. Previous naive document processors extracted text without concern for visual organization that is obvious to a human observer, making downstream extraction unnecessarily difficult. Using VILA can reduce the error rate of large language model processing by roughly 10%.

2. **Embedding recycling** is a method for accelerating expensive neural model training in the very common setting where a model is repeatedly retrained on an evolving document collection. A large amount of the training procedure exists simply to move parameters into the right rough "neighborhood" for handling the task at hand. This method caches sets of parameters from previous iterations of a trained model, allowing later training attempts to exploit past work without starting

from scratch. On two important tasks for scientific documents — Named Entity Recognition and classification — this method can yield a 55-86% training speedup with only a 0.2% loss in F1 when compared to the non-recycled setting.

3. **Affiliations Linking** is a module for deduplicating organization names that often appear in text with small permutations. For example, a human can tell that the MITt chemical engineering dept, Cambridge, MA, USA, usa refers to the same real-world entity as Massachusetts Institute of Technology (MIT) but not to Massachusetts Chemical Engineering LLC. This problem is endemic in information extraction, and especially problematic when integrating huge numbers of texts into a single resource, as with an OKN. By implementing a pre-ML filtering step that removes obviously improbable candidates, this method is able to obtain linking results that are equal to a conventional Transformer model, at less than half the computational cost.

4. **CascadER Knowledge Graph Link Prediction:** The CascadER system aims to make an ambitious OKN goal — that of adding novel scientific knowledge to an OKN — possible. If we could find the set of potential links in an OKN that are supported using published evidence, it would amount to creating a structured representation of scientific knowledge. Methods exist to do this, both purely from text and (more productively) from multimodal inputs such as text plus figures.

   Unfortunately, state-of-the-art approaches take up to 11 days of processing to evaluate a single novel OKN edge (because they test every possible answer using the entire text as evidence). The CascadER system is structured to evaluate evidence in a particular order, from "coarse and cheap" to "fine and costly". By better using computational resources, it can obtain substantially better quality results than competing methods.

Knowledge Network Provenance System

The Knowledge Network Provenance System (KNPS) software system enables easier data production for OKNs. Provenance systems attempt to capture a record of data operations for later use. If we had a record of all OKN-relevant data processing, we could answer a range of questions that would greatly accelerate development, including:

- What training set was used to generate a particular model?
- Is a dataset suitable for use with medical products?
- Was a particular visualization created by people I trust?

Unfortunately, provenance systems are not widespread enough to capture this information in most circumstances today. Traditional provenance systems — sometimes seen in financial applications, or certain relational databases, or large-scale ML platforms — are characterized by (a) substantial software

modifications and (b) detailed and precise operations of each data record. Point (a) means that collecting provenance information is extremely expensive. Since modern OKN and other data production pipelines often involve a range of teams, datasets, tools, and institutions, even substantial investment in provenance software can be insufficient for capturing the end-to-end information necessary to answer the questions above.

We note that the cost advantages of social OKN dataset construction might allow us to address this problem. OKNs can yield datasets that are much larger and cheaper than traditionally administered databases. They do this by combining machine learning and social mechanisms to make it feasible to ingest large sets of data from a wide variety of sources. The resulting dataset may not be as "perfect" as a traditional database but is often much larger with quite good quality.

Our KNPS provenance system follows this approach, by:

1) Admitting low-quality instrumentation information from desktop clients, cloud service logs, and other scanners
2) Using machine learning and social OKN mechanisms to upgrade this cheaply collected provenance data when appropriate, and to infer missing data when possible.

The current platform has a range of scanners, a graphical front-end for examining the collected data, and a set of mechanisms for inferring and upgrading the low-quality inputs.

CORD-19 is a large OKN containing scientific information about the coronavirus, extracted from more than 1M papers. It was produced using the extraction software described above.

Videos, websites, and other resources

The following resources provide additional information about the KNPS.

- The CORD-19 dataset: **GitHub - allenai/cord19: Get started with CORD-19**
- The KNPS system: **GitHub - mikecafarella/KNP**
- The suite of extraction tools have been used to create a public-facing repository of scientific paper extractions: **Semantic Scholar.**

Track integration goals and accomplishments

Our primary in-track integration work is with the SCALES team. This work is described above in the SCALES section. The primary work from the KNPS perspective lies in creating provenance recordings that can be repurposed for future configuration and deployment. Technically, there exists an interesting opportunity similar to that of large-scale text "foundation models" like GPT-3 or GitHub Copilot that can capture many tasks in a single trained model.

We aim to collect a large number of provenance recordings, first for SCALES, and then for other tools as well. The resulting provenance OKN can be used to construct a model that describes the overall distribution of software configurations, much in the same way that large-scale text models can capture the overall distribution of useful natural language sentences. At later use time, an administrator should simply specify a tiny number of distinctive configuration parameters, and then ask the model to fill in the remaining information.

We have started to collect provenance information, and have ancillary code ready for constructing a trained embedding-style model from OKN information. Once data collection is complete, it should be fairly straightforward to test how accurately we can autoconfigure Satyrn with this statistical record of past usage.

National Science Foundation
Directorate for Technology, Innovation and Partnerships

BETA.NSF.GOV/TIP/LATEST