Representing burn assessments in knowledge graphs

Brian Wee, Ph.D.
Massive Connections, LLC
Washington, D.C.
(retired)

Connecting
Data,
Science,
Technology,
Policy to
Societal Needs

Context setting #1 of 3: preamble

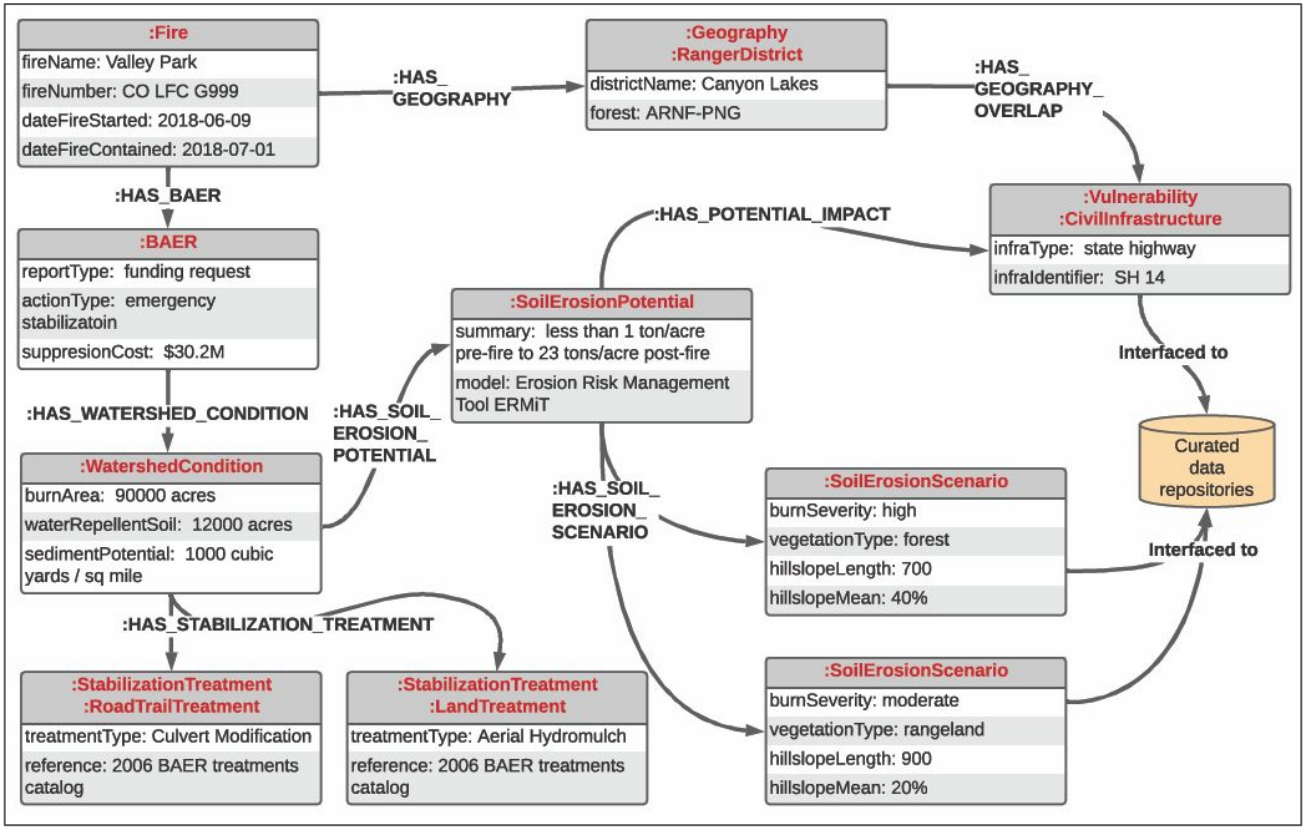
- Outcomes of experiments performed circa 2020 / 2021 under the auspices of:

ESIP = **E**arth **S**cience **I**nformation **P**artners

- For the sake of brevity, imprecise vocabulary (i.e. at times technically incorrect terminology) will be used in this presentation.

Context setting #2 of 3: Key take-away

Knowledge graphs (KGs) can be used to enable wildfire data and information interoperability across the fire-management lifecycle. KGs assert facts, and are therefore *likely* useful for reducing AI hallucinatory incidents in large language models (e.g. GPT-3.5 that powers ChatGPT).



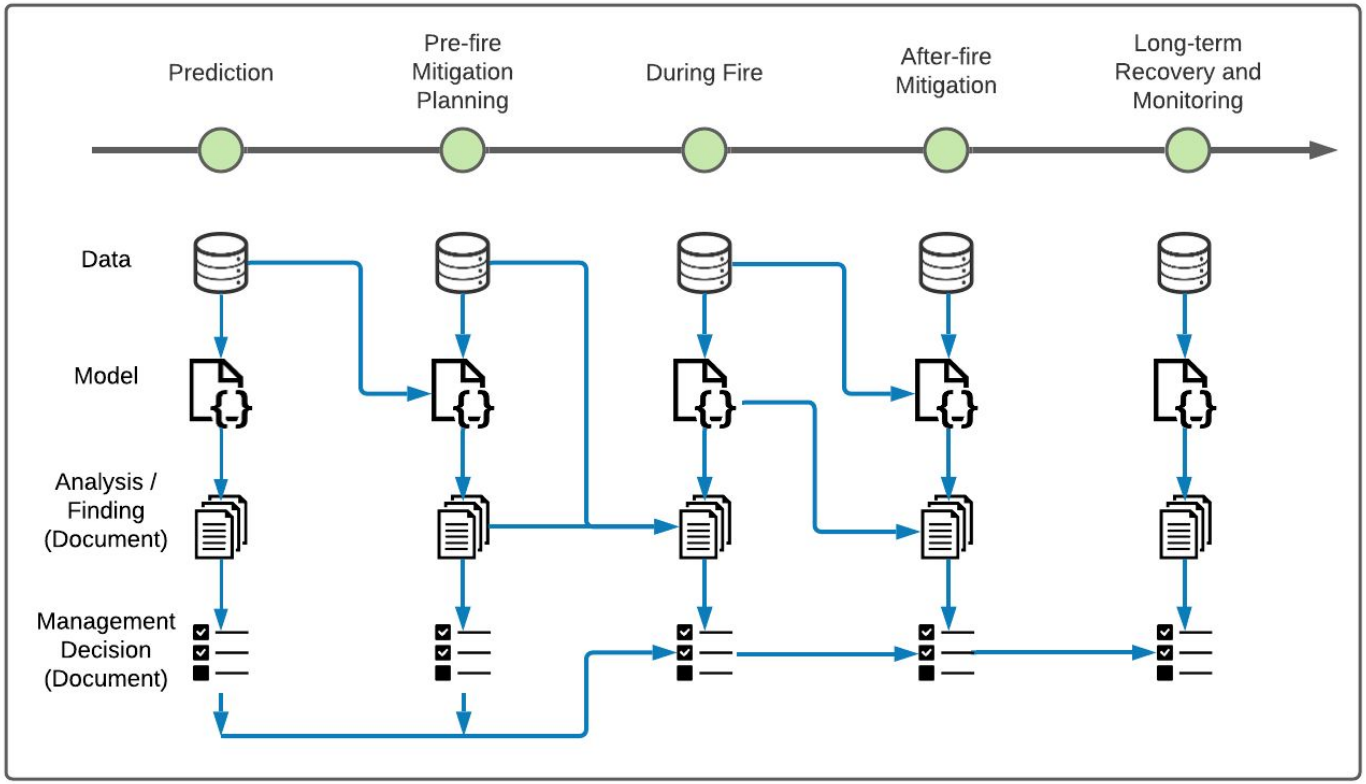
Context setting #3 of 3: Your mental frame for this talk

We shall be focusing on the **bolded areas** below.

- Fire management life-cycle phases (roughly):
  - (1) Prediction, (2) Pre-fire mitigation planning, (3) Fire fighting
  - **Mitigation assessment and planning after fire containment*****
  - (4) After-fire mitigation, (5) Long-term recovery and monitoring
- Structured versus **unstructured** data and **information**.
- AI methods may be very roughly conceptualized into:
  - Statistical methods (e.g. support vector machines)
  - **Symbolic / knowledge representation methods (e.g. ontology)**
  - Artificial Neural Network methods (e.g. deep learning)

*** *Not really considered a phase alongside the other five phases, but represented as a phase for presentation purposes.*

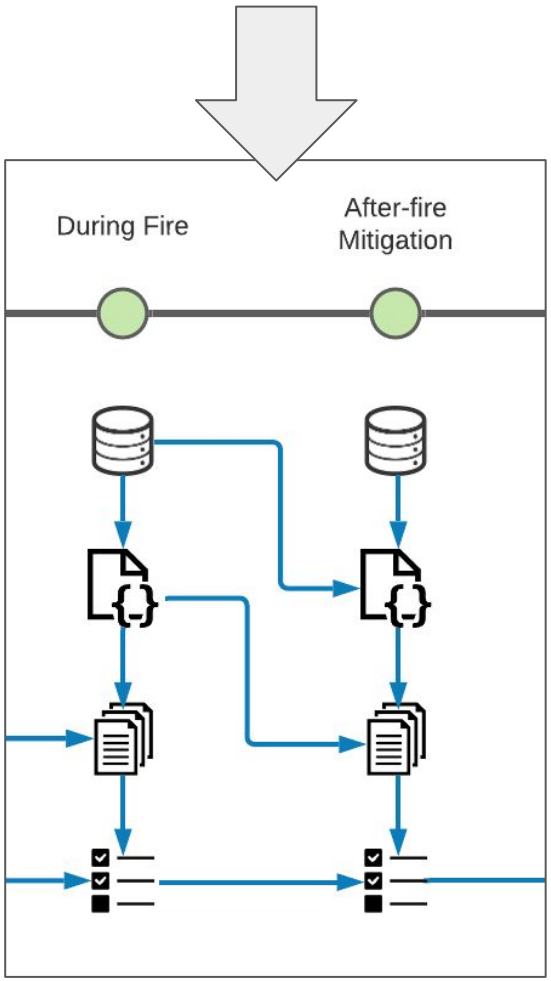
Reusing Data and Information Across the Fire Management Lifecycle

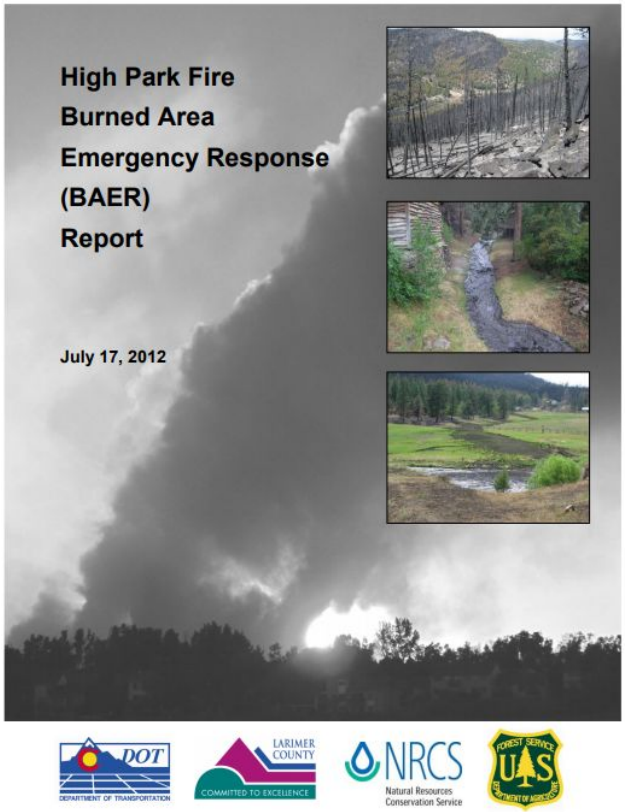


Wee, Brian; Teng, William; Zader, Dave (2021): Wildfire data and information interoperability across fire management phases. ESIP. Poster. https://doi.org/10.6084/m9.figshare.14974266.v1

Focus for this talk

Right after fire containment





- BAER (pronounced “bear”) report used to request and justify funds for after-fire mitigation measures (e.g. soil erosion prevention).
- Reports tend to contain certain types of information.

Typical contents of BAER reports

(Highly abstracted)

- Burned area description
  - Topography
  - Vegetation types
  - Dominant soils
  - Burn severity
  - Soil repellency to water
  - Soil erosion characteristics
- Hydrologic characteristics
  - Assumptions about future precipitation events
  - Estimated water discharge parameters before and after fire
- Risk assessment
  - Values at risk (e.g. physical infrastructure, biodiversity, ecosystem services, cultural values)
  - Options for emergency treatment (e.g. mulch application, culvert enlargement, concrete barriers)

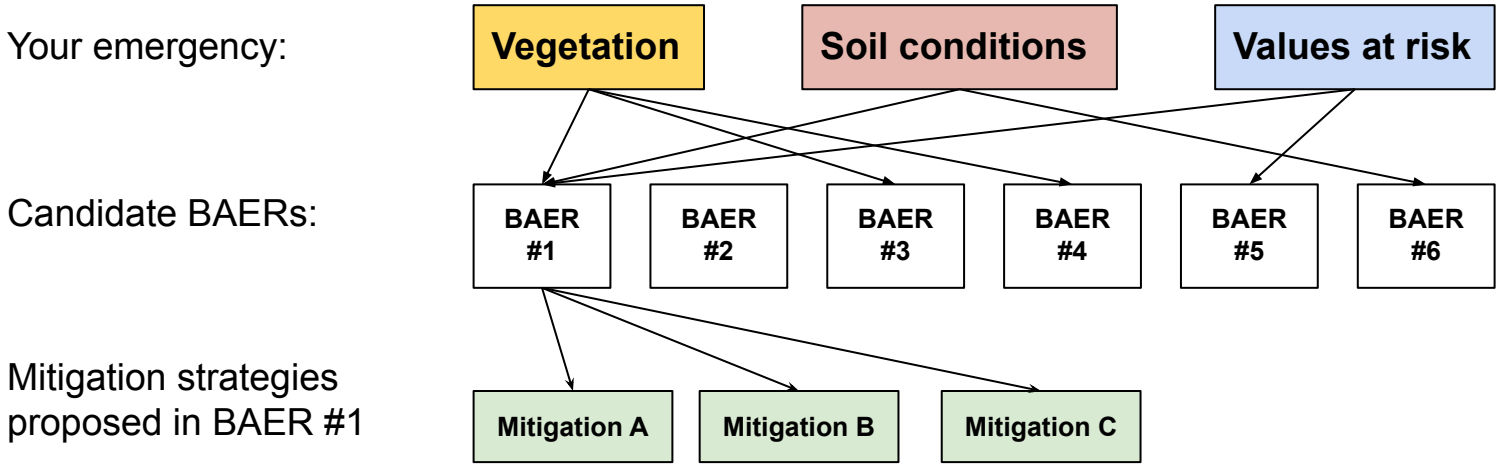
Question: Can we use technology to mine the unstructured information from past BAER reports for a current emergency?

Scenario

- BAER team uses natural language to describe the current emergency in terms of burn area, hydrologic conditions, and values at risk. E.g.

> *Osprey Creek Fire.. approximately 87,487 acres… northern foothills of Larimer County... predominantly mature lodgepole pine, relatively closed canopy... igneous formation of granite rock material cover 19 percent of the area… Erosion potential within the burn ranged from less than 1 ton/acre pre-fire to 23 tons/acre post-fire… burn area is adjacent to Native American culturally important site with critically endangered flora… burn area is also adjacent to army training grounds that may contain unexploded ammunition.*

- Algorithm matches current emergency to old BAERs and justifies ("explains") its recommendations in a manner like:



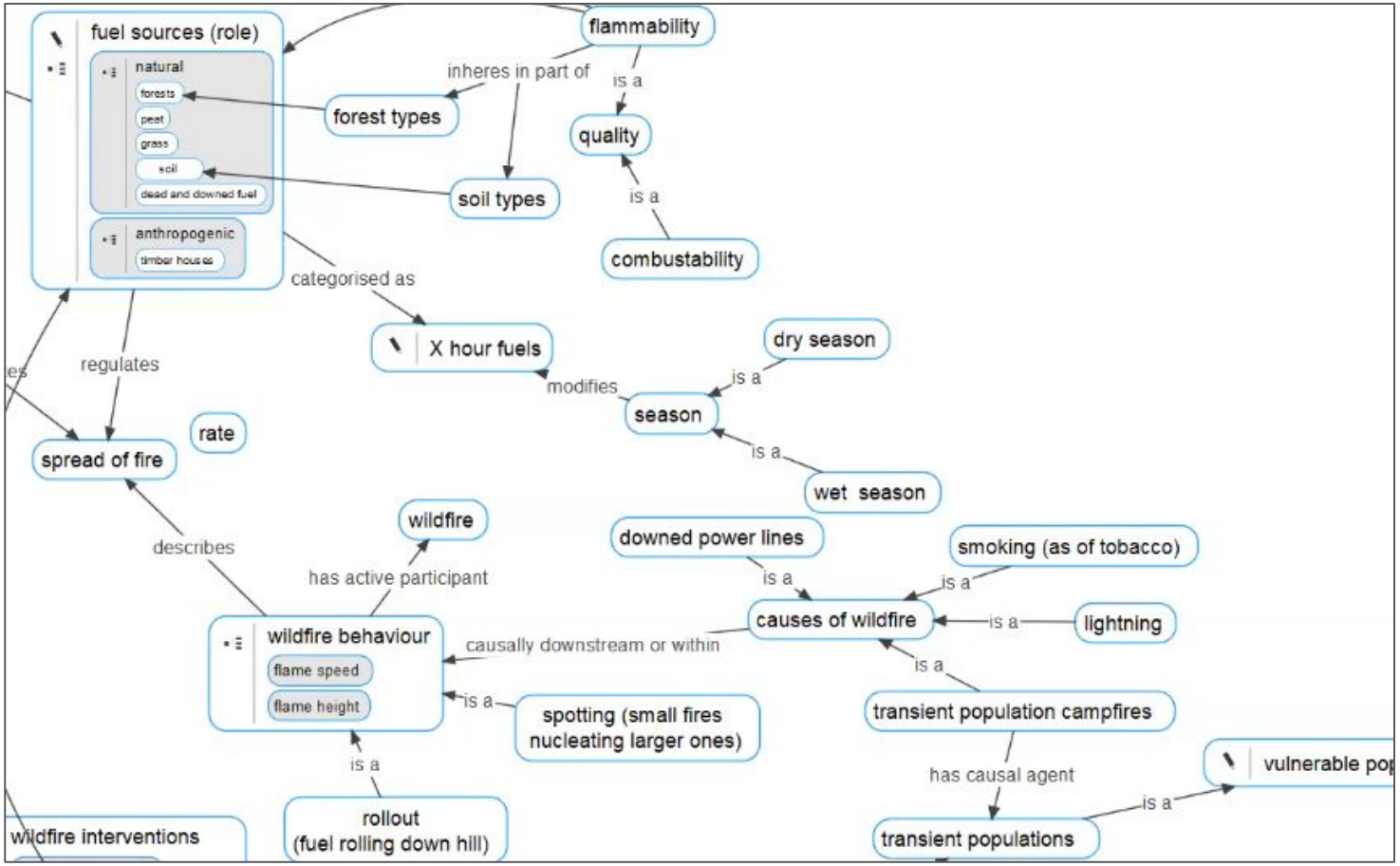
Technology approaches

- **Artificial Neural Networks.** Challenges with contemporary large-language models include:
  - Lack of explainability: “show me how you arrived at your recommendations”
  - Hallucinations: factually incorrect answers (probabilities vs. assertions)

- **Symbolic / knowledge representation.** Challenges include:
  - Human-resource intensive: requires experienced knowledge engineer to elicit knowledge from subject matter experts (SMEs: e.g. wildland fire risk assessors, soil scientists, hydrologists, meteorologists).
  - Maintenance of knowledge graph.

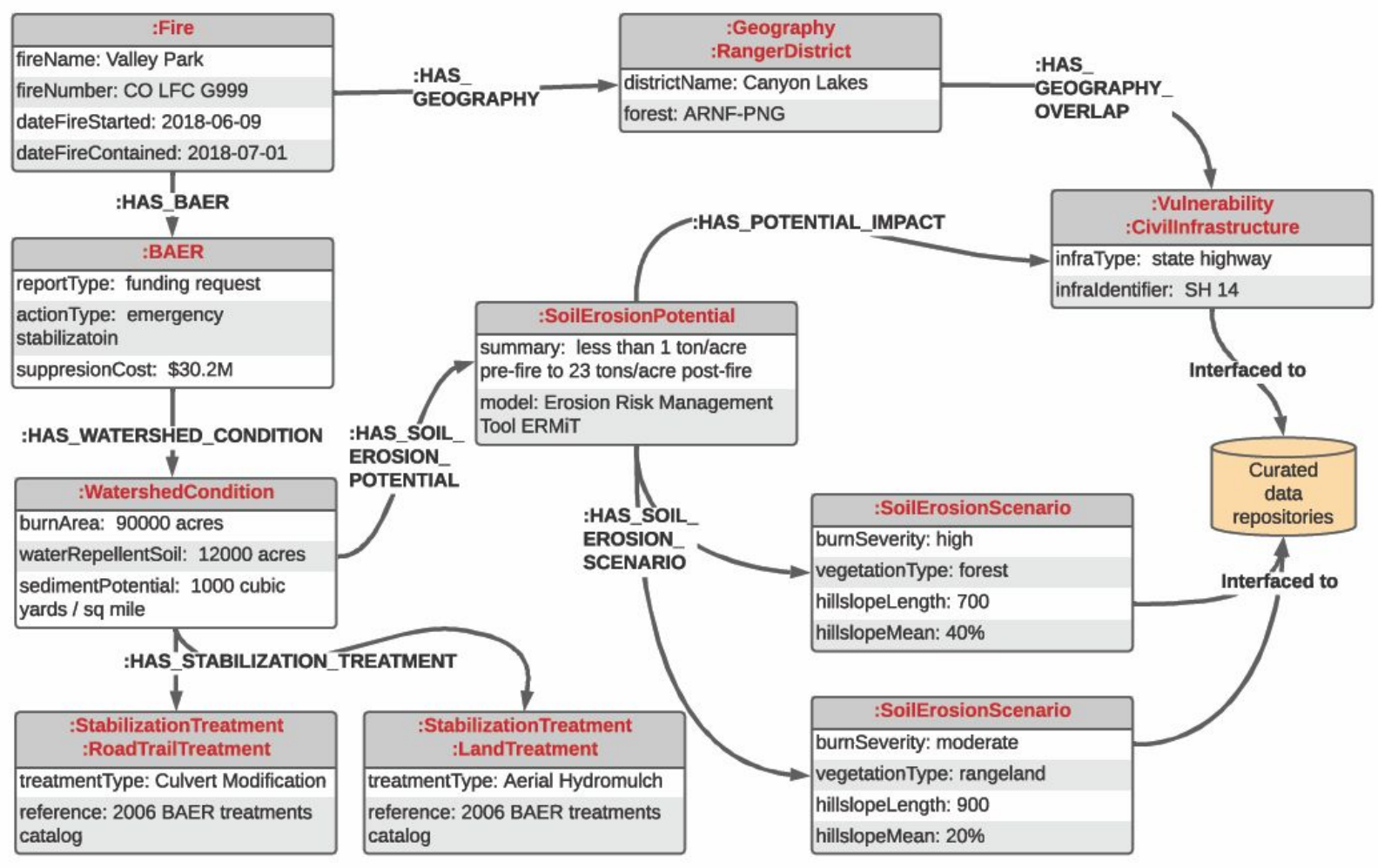
Ultimately, a **hybrid approach is probably the answer**, but work in hybrid approaches are very nascent (~3 years old).

Knowledge elicitation with SMEs



Screenshot of knowledge elicitation and capture during a July 2021 ESIP workshop by Pier Buttigieg (Helmholtz Metadata Collaboration)

Wildfire BAER knowledge represented (asserted) as a graph

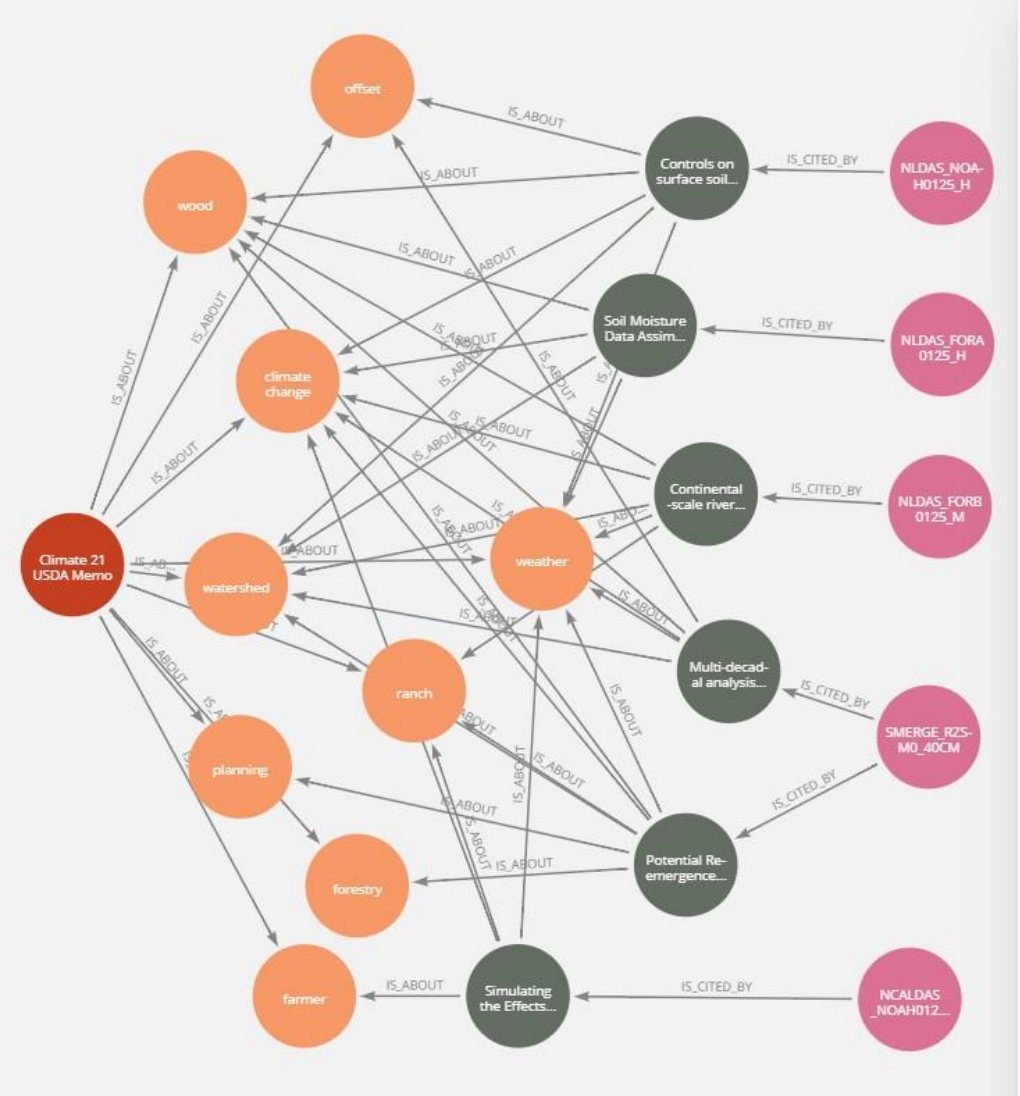


Hypothetical representation of fire information extracted from reports ingested into a Neo4j graph database.

Automated parsing of documents to a graph database

Python script for populating a Neo4j graph database with relationships between between data, publications, shared-concepts (e.g. “watershed”, “forestry”, “offset”), and policy recommendations regarding carbon management for the (then) incoming Biden Administration.

(ESIP experiment conducted in Jan. 2021 using Neo4j.)



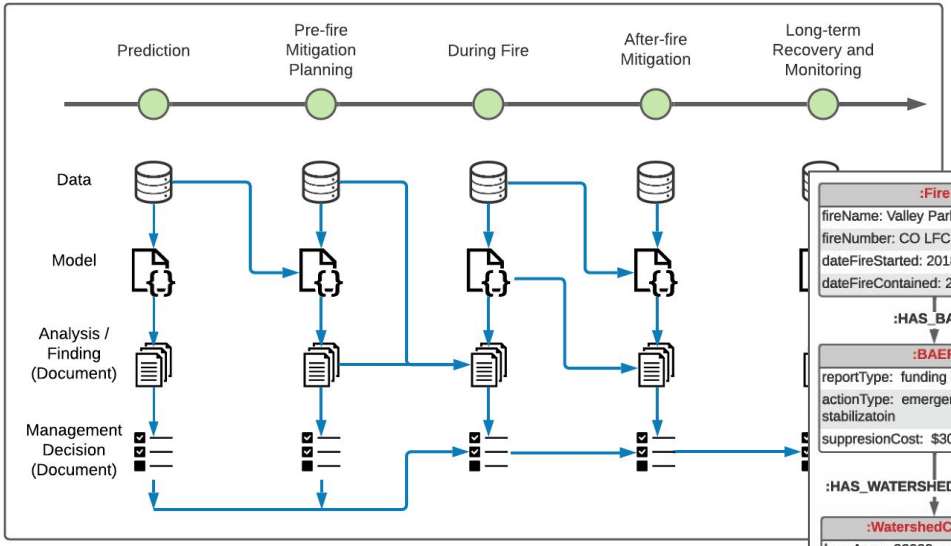
Trajectory for managing and using wildfire knowledge

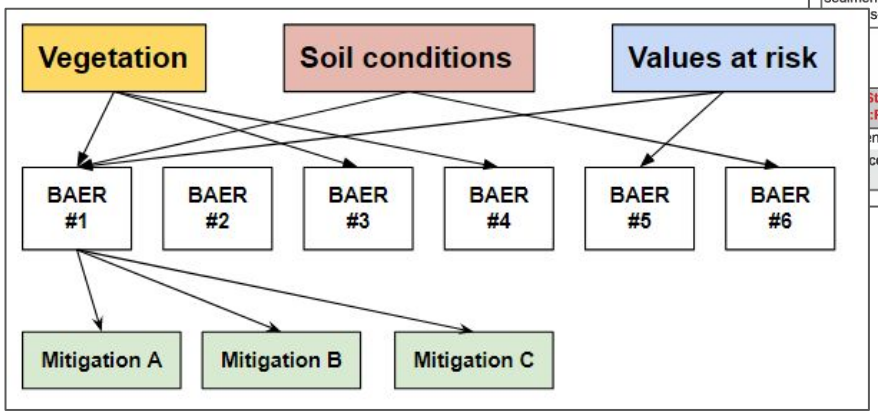
- Discussed in ESIP meetings on using KGs to:
  - **Coordinate wildfire resilience efforts** across different organizations / projects by identifying policies, geographical data, priority values at risk, etc. that are shared between different communities.
  - Share **strategies for using data and models to inform decisions** by explicitly linking wildfire prevention plans to scientific publications that are linked to data and models.

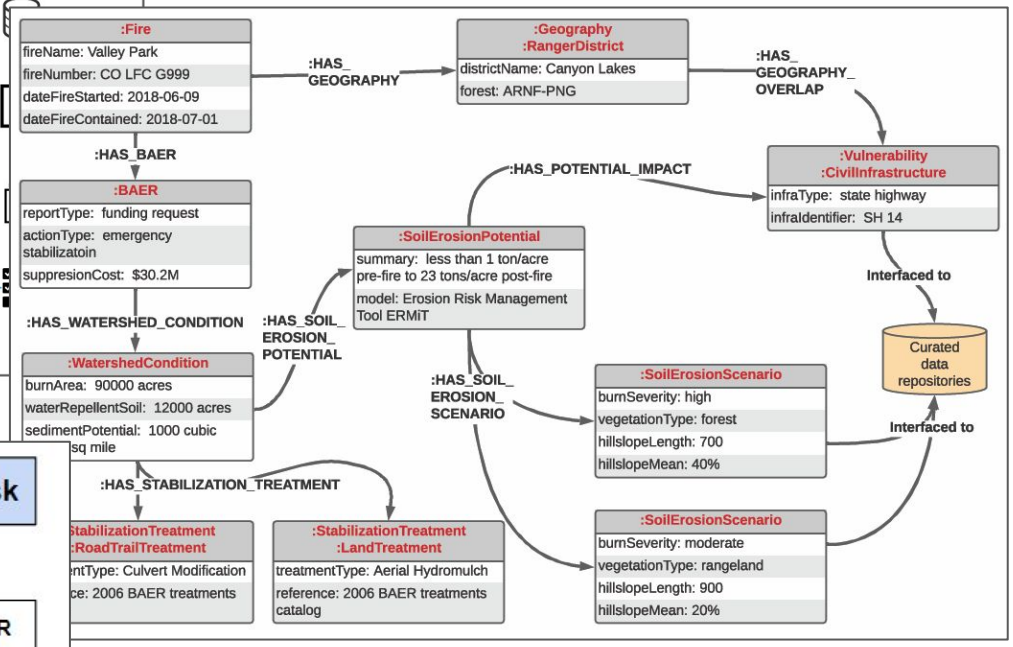
- Large language model (LLM) integration with knowledge graphs (KG):
  - LLM translates the user's natural language question into graph database query language.
  - Factual assertions in KGs are used to **train LLMs** (very recent active research, e.g. Google Research and other private sector entities).

Eventually: Use hybrid LLM+KG AIs to produce draft BAER reports that link to publications, data, and models (?).

Visual recap







Brian Wee, bwee@massiveconnections.com, Washington, D.C.

Massive Connections

Connecting Data, Science, Technology, Policy to Societal Needs