

Tools for the Citation and Analysis of Heterogeneous Data in Critical Infrastructure

Gabriel A. Weaver

University of Illinois at Urbana-Champaign

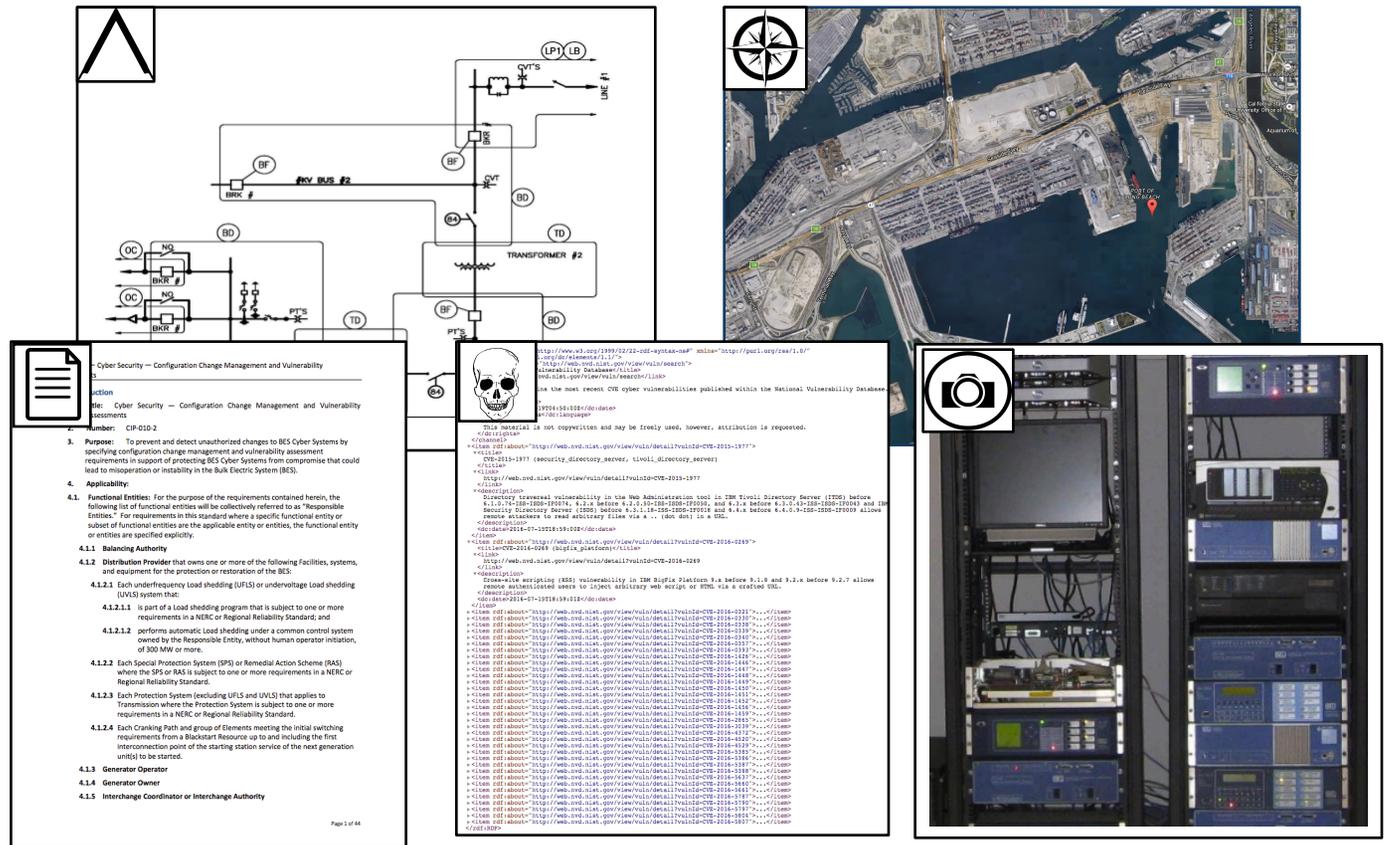
gweaver@illinois.edu

We want to enable **efficient communication of information about critical infrastructures** among government, industrial, and academic communities.

Stakeholder		Information Sharing and Analysis Decision Needs	
		<i>Strategic</i>	<i>Tactical</i>
Government		Executive Order (EO) 13636, Presidential Policy Directive (PPD) 21 NIST Cybersecurity Framework, Funding and Acquisitions	Disaster recovery and response (DHS)
Industry	Energy	2011 DOE Roadmap, NERC-CIP Regulations, Planning, Contingency Analysis	Situational Awareness for Operations
	Maritime	GAO-14-459, Acquisitions, Port All-Hazards Planning	Situational Awareness, Equipment Acquisition
	Manufacturing	Digital Thread, Supply Chain Analysis, Acquisitions	Tolerance Information
Academia		NSF CIF 21, ARPA-E GRIDDATA	Reproducibility of prior work for research, availability and utility of prior datasets and analyses

Stakeholders must share and analyze heterogeneous data from various locations.

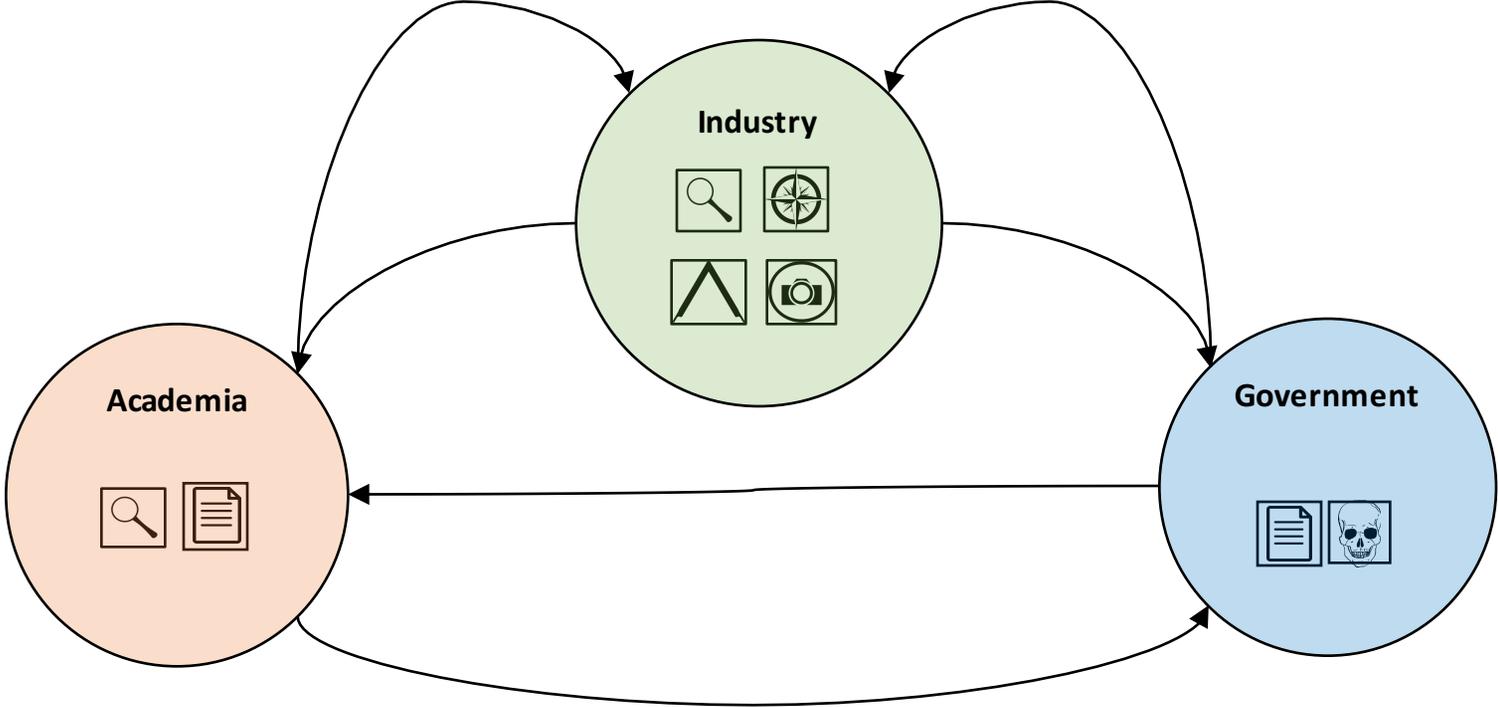
-  Documents
-  Geographic
-  Tool Outputs
-  Schematic
-  Video
-  Vulnerabilities



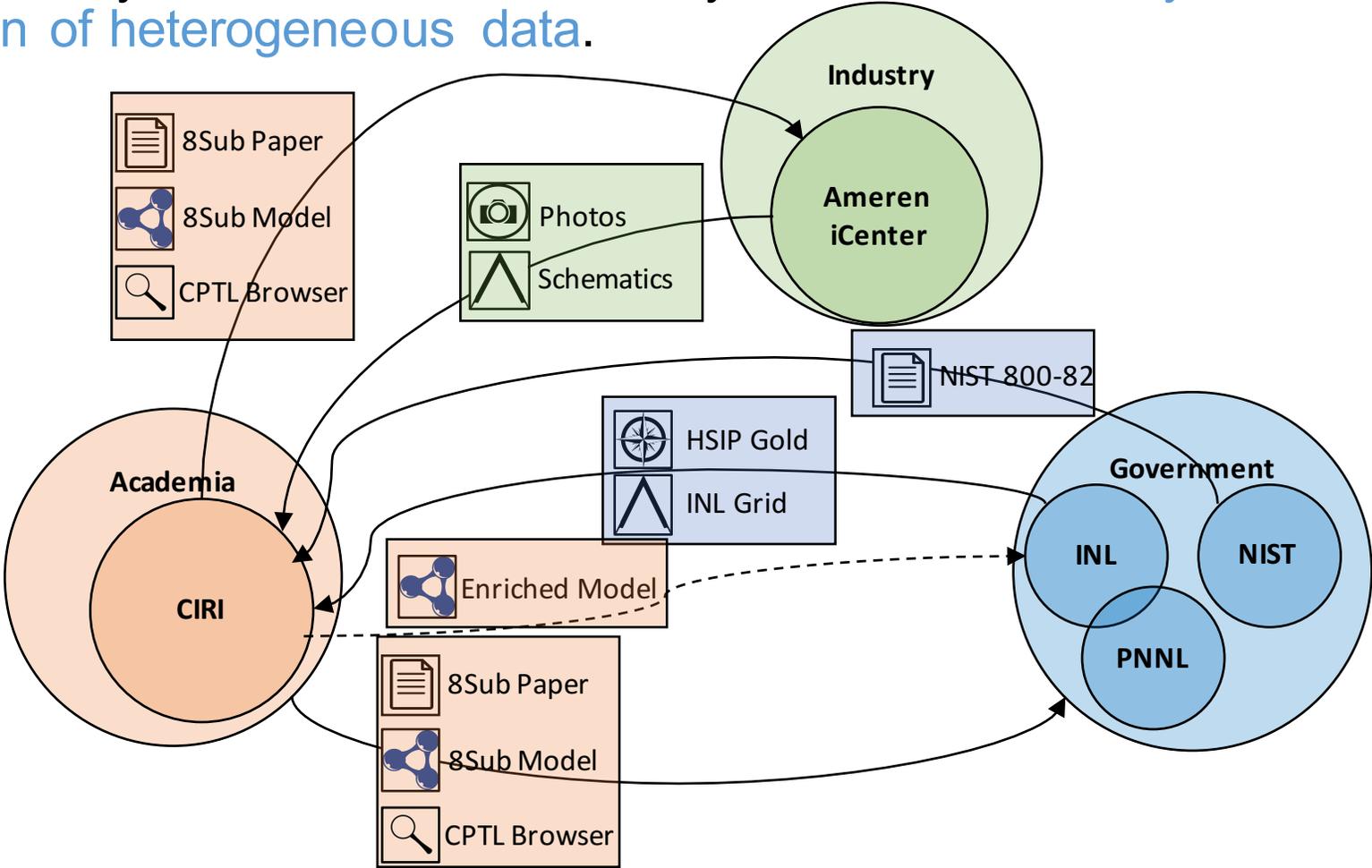
The image displays a collage of heterogeneous data types used in system analysis:

- Schematic:** A detailed power system diagram showing components like transformers, buses, and protection devices.
- Geographic:** A satellite map of an industrial or urban area with a red location pin.
- Documents:** A snippet of a document titled "Cyber Security - Configuration Change Management and Vulnerability Assessments" with numbered sections (3, 4, 4.1, 4.1.1, 4.1.2, 4.1.2.1, 4.1.2.2, 4.1.2.3, 4.1.2.4, 4.1.3, 4.1.4, 4.1.5).
- Vulnerabilities:** A screenshot of a vulnerability scan report showing a list of CVEs (e.g., CVE-2013-1977, CVE-2014-0228) and their associated URLs.
- Video:** A photograph of server racks in a data center.

A **wide variety of stakeholders** need to share information on Critical Infrastructure Cybersecurity.

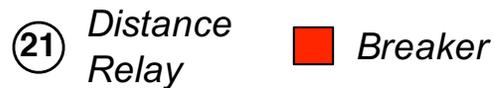


Social interactions of our research set the stage for a socio-technical system to evaluate security claims via a theory for citation of heterogeneous data.

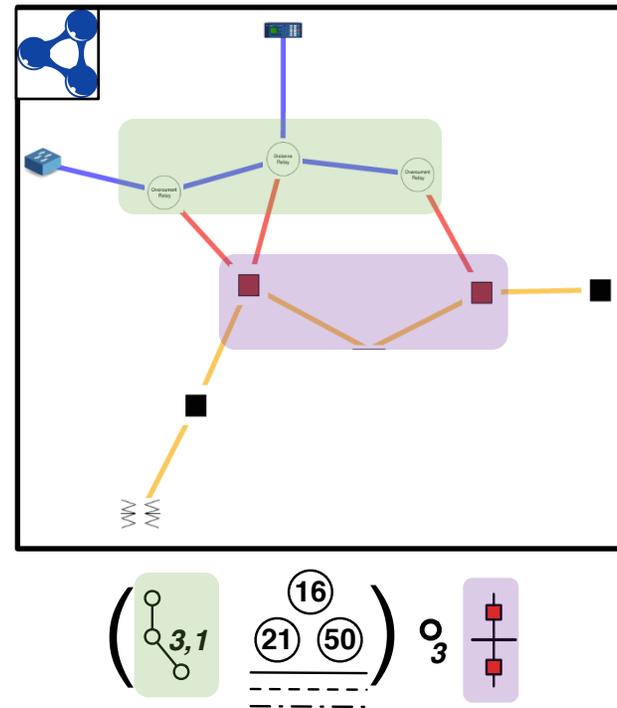
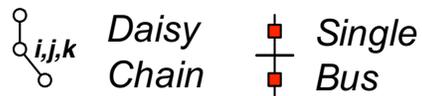


In order to make a security claim about a system, stakeholders need to be able to inventory assets and dependencies.

1. Ontologies define assets and their interactions.

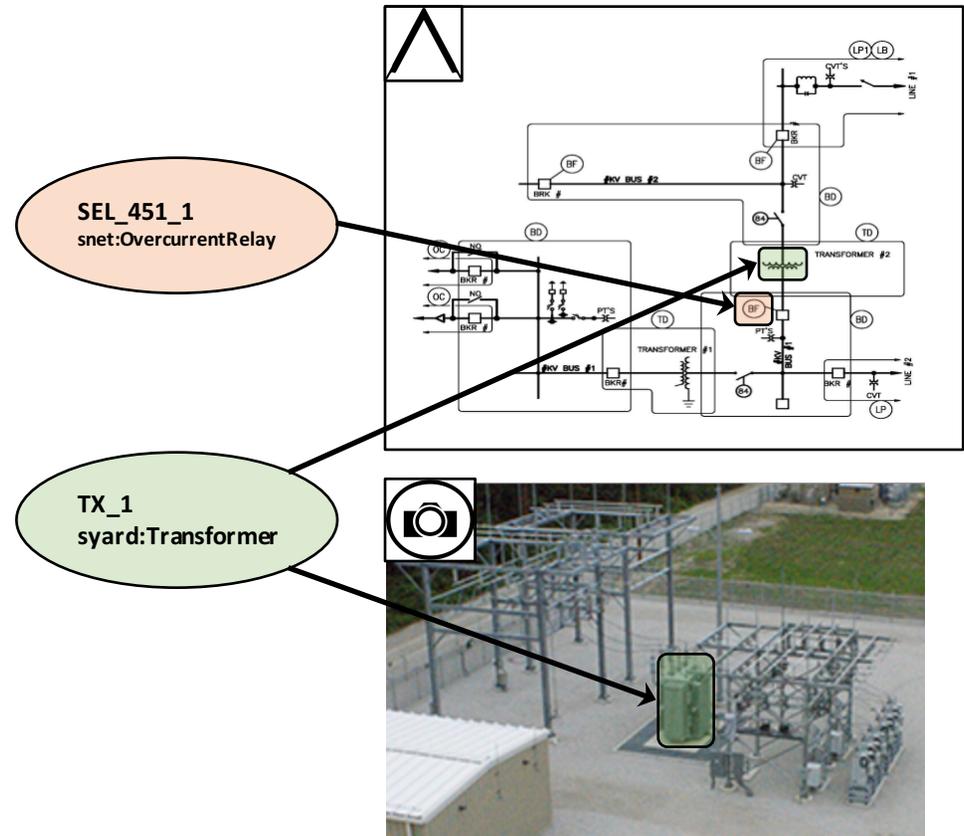


2. Templates encode and recognize higher-level structures.



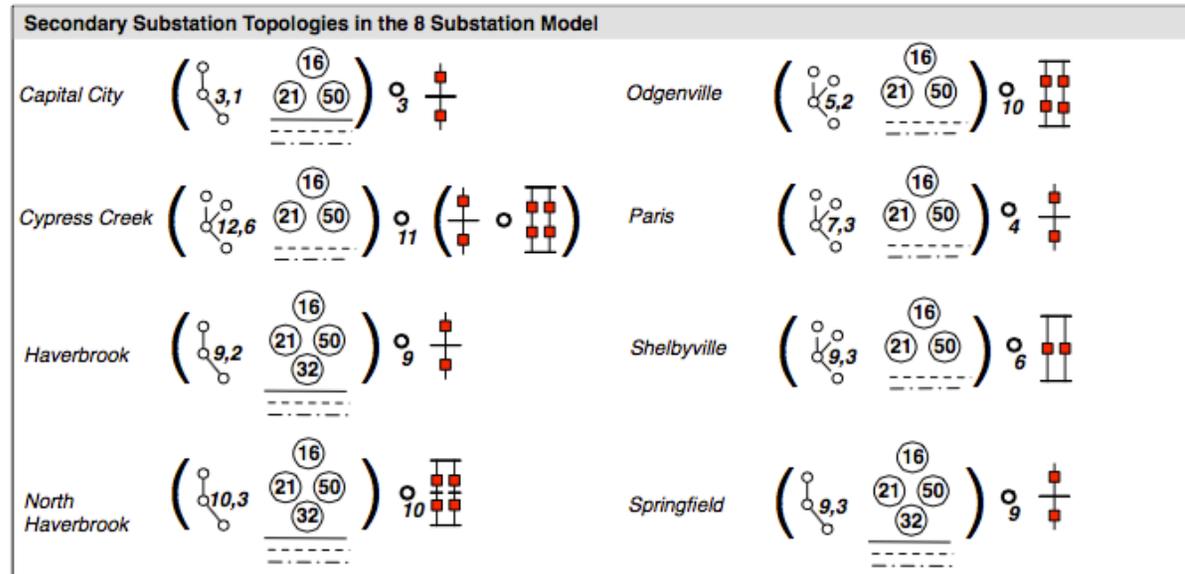
Practitioners need a **controlled vocabulary** to describe assets and their interactions.

- Need to define **types of entities** to align expressions in heterogeneous data.
- Define **ontologies** to represent types of assets and interactions.
 - Control center network
 - Substation network
 - Substation yard
- Ontologies available at <http://cptlc.github.io/>



We have created an 8-substation model using CPTL so academia has access to artificial but realistic datasets.

- Example of how to encode cyber-physical topologies
- Benefit researchers with an artificial but realistic dataset
 - Compare analyses
 - Develop metrics
- To appear at IEEE SmartGridComm 2016



8 Substation Model Browser

Substation Inventory

- 8 Substation Model
 - Control Center
 - Substations
 - Capital City
 - Capital City Network
 - Capital City Yard
 - Cypress Creek
 - Cypress Creek Network
 - Cypress Creek Yard
 - Haverbrook
 - Haverbrook Network
 - Haverbrook Yard
 - North Haverbrook
 - North Haverbrook Network
 - North Haverbrook Yard
 - Odenville
 - Odenville Network
 - Odenville Yard
 - Paris
 - Paris Network
 - Paris Yard
 - Shelbyville
 - Shelbyville Network
 - Shelbyville Yard
 - Springfield
 - Springfield Network
 - Springfield Yard

Asset Connectivity

The diagram illustrates the connectivity between various assets in a substation. It features several nodes: a black square at the top left, a red square in the upper middle, a blue square on the right, a black square at the bottom right, and a transformer symbol at the bottom right. A central node is connected to multiple other nodes via colored links: orange, green, blue, and black. The links represent different types of connections between the assets.

Graph Style

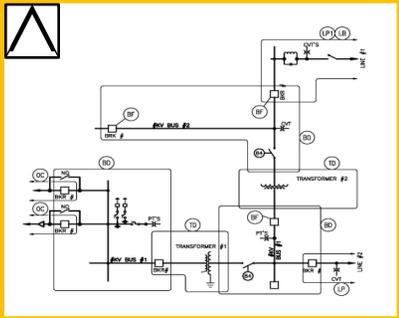
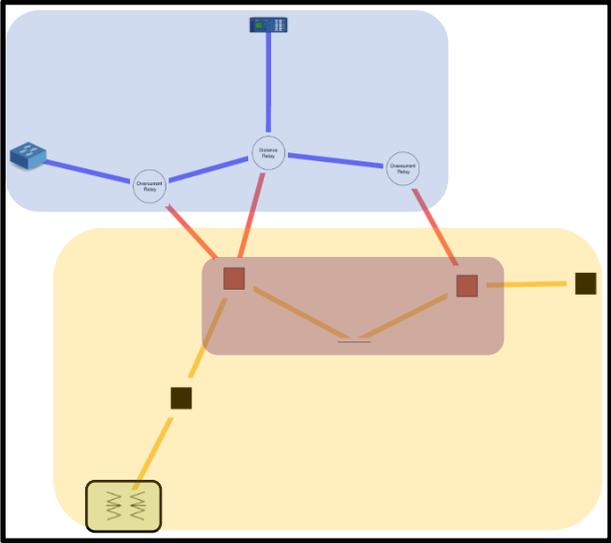
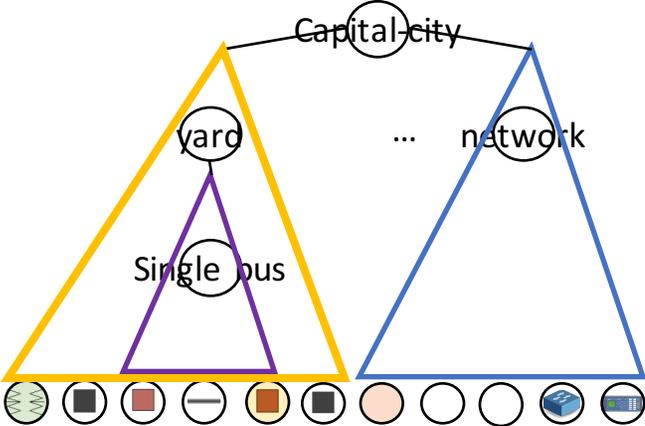
Style Nodes

Node Type: Breaker
Color: black
Size: medium
Apply

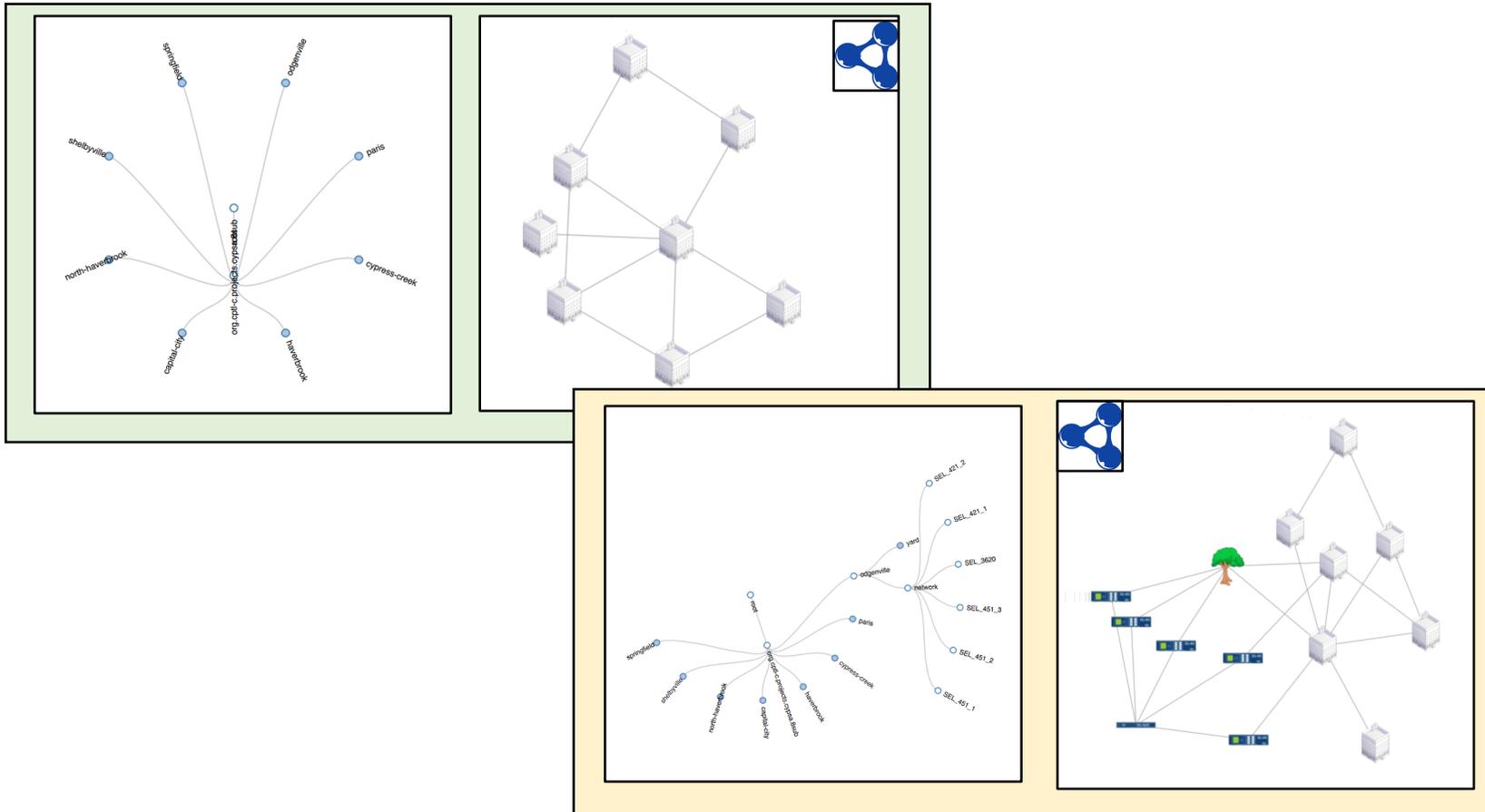
Style Links

Link Type: Serial Link
Color: blue
Width: medium
Apply

The hierarchical order enables them to cite and navigate information at multiple levels of abstraction.



Hierarchical Model Browser



Schematic/Image Browser

Image



An aerial photograph of a power substation. In the foreground, there is a small, tan-colored building with a white roof. To the right of the building, there is a large, open area with a concrete or gravel surface. In the background, there are several tall metal structures, likely part of the power distribution system, and a large red cylindrical tank. A green sign is visible near the structures. The substation is surrounded by a fence and trees. A smaller inset image in the top right corner shows a close-up of a distribution voltage regulator with navigation controls (back, forward, and zoom).

Asset Style

Style Nodes

One Line Node Type:

Color:

Size:

Style Links

Link Type:

Color:

Width:

Ongoing and Future work

1. Incorporate **additional data sources**
 - Network Traffic: NMap, PCAP
 - Streaming data
 - Schematic Information
2. Define more **entities and templates for Critical Infrastructure**
 - Expand ontologies to include IEC 61850, C37.22, PSADD
 - Expand templates to include NIST 800-82
3. Create more **analyses based on citation structure**
 - Investigate multiple citation structures for same datasets
 - Automated redaction via ontologies
 - Comparison over time of heterogeneous datasets
 - Multimodal security policies

Conclusions

1. We want to enable **efficient communication of information about critical infrastructures** among government, industrial, and academic communities
2. The **purpose of information sharing** and analysis is to **communicate arguments** about the security of a system
3. Current modes to communicate such arguments---such as our **publication system---**
cannot keep pace with the systems we study
4. Therefore, we **rethink the communication of scientific argument** and encode it as an **explicit data structure** whose components may be shared and composed.
 - Data Sources
 - Citation Scheme
 - Analyses and Visualization
5. Our goal is to design and implement an ecosystem to **accelerate and gain telemetry on the research process** while simultaneously increasing the ability to **compose/compare research results**, and **transfer them to practice within industry**.
6. **CIRI is uniquely positioned to create such a socio-technical system** for the discipline of Critical Infrastructure System Security.