Knowledge Framework for Critical Infrastructure Dependency Analysis

Towards a All Hazards Vulnerability and Risk Assessment Capability

5/19/2015

www.inl.gov

Idaho National Laboratory

Agenda

- Introduction
- All Hazards Knowledge Framework (AHA)
- Technical Approach
- Preliminary Results

Critical Infrastructure Challenge

Interconnected

Climate

Aging

Cyber

Disparate

Dependent

Physical

Vast

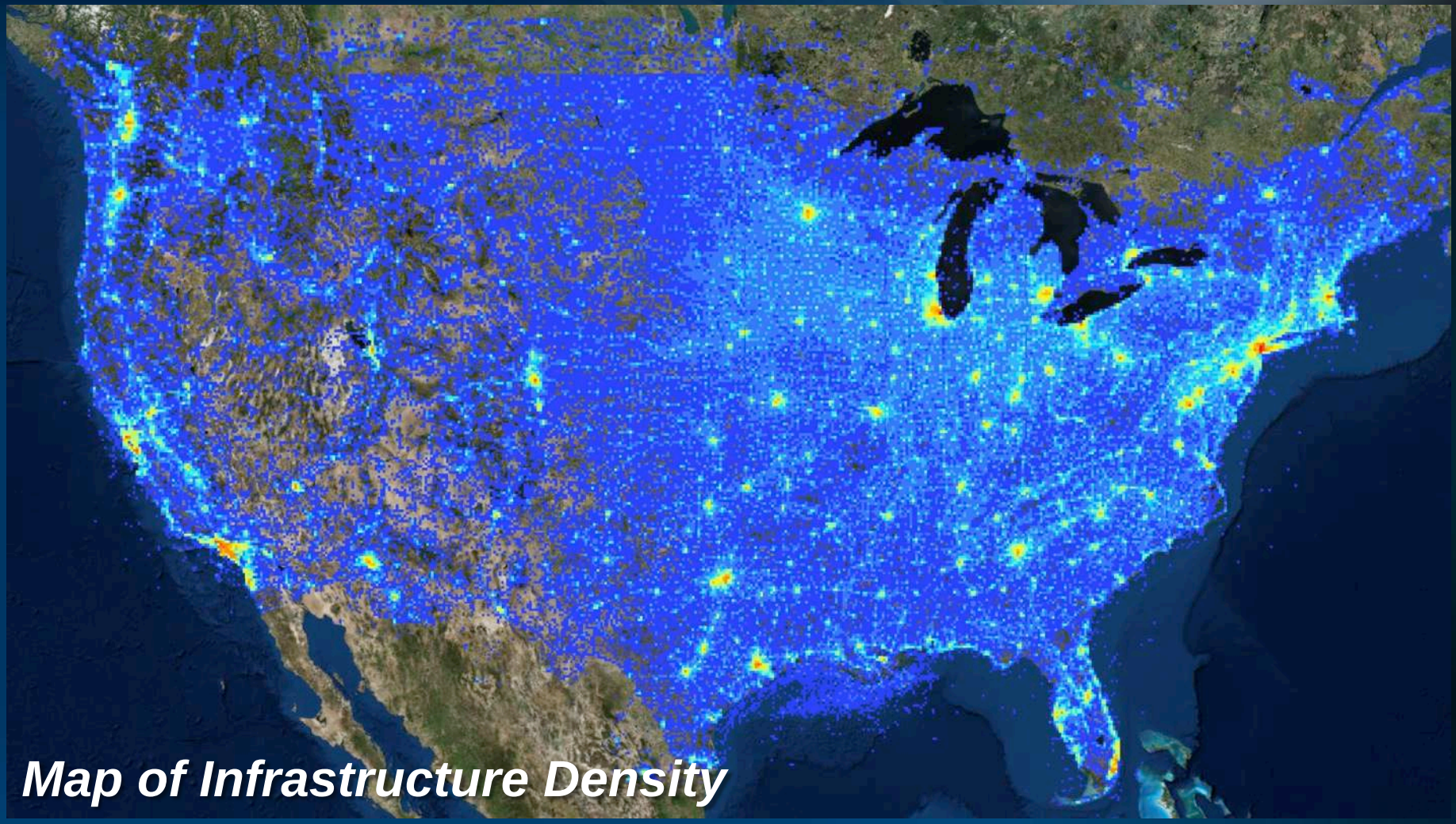
Lifeline

Resilient

Hazards

Dynamic

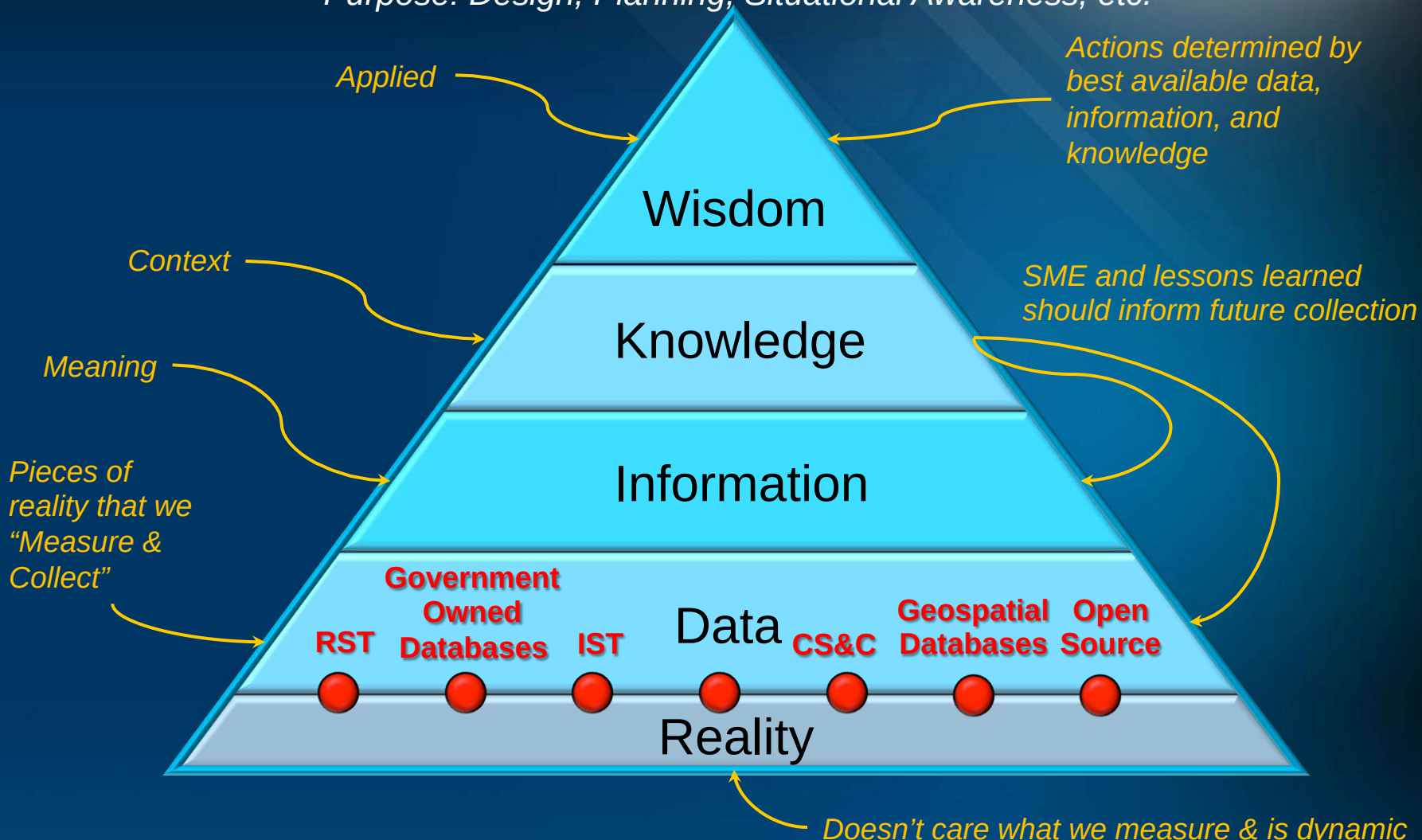
Current Situation



Map of Infrastructure Density

Data to Action

Purpose: Design, Planning, Situational Awareness, etc.



Enable risk and prioritization decisions

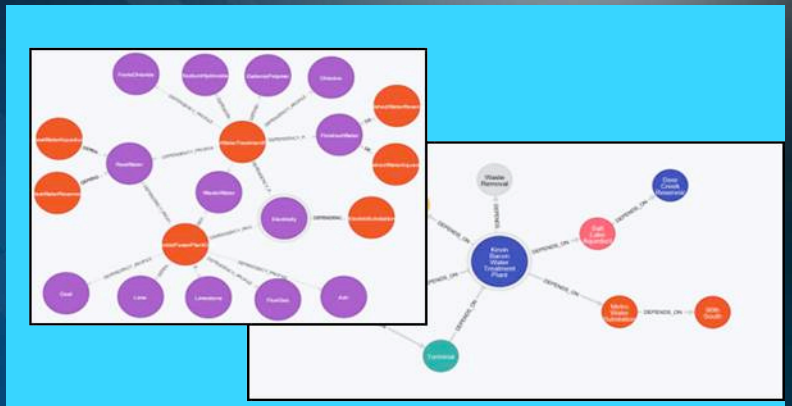
- Actionable information – Getting the right information to the right person at the right time, and in a form they can efficiently use

  - Uncover trends associated with failures
  - Discover Facilities
  - Document Dependencies

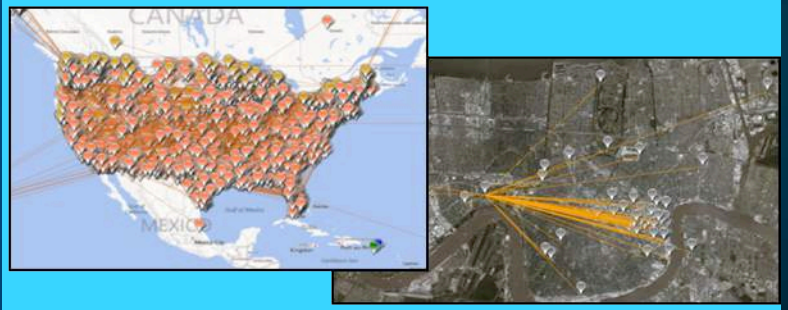
Prepare | Protect | Mitigate | Respond | Recover

What is AHA?

- A framework that learns from data and expert knowledge
- Seamlessly integrates structured and unstructured datasets
- Provides a more holistic dataset for analysis
- Provides both geospatial and graph visualization capability due to problem space complexity



Dependency Model



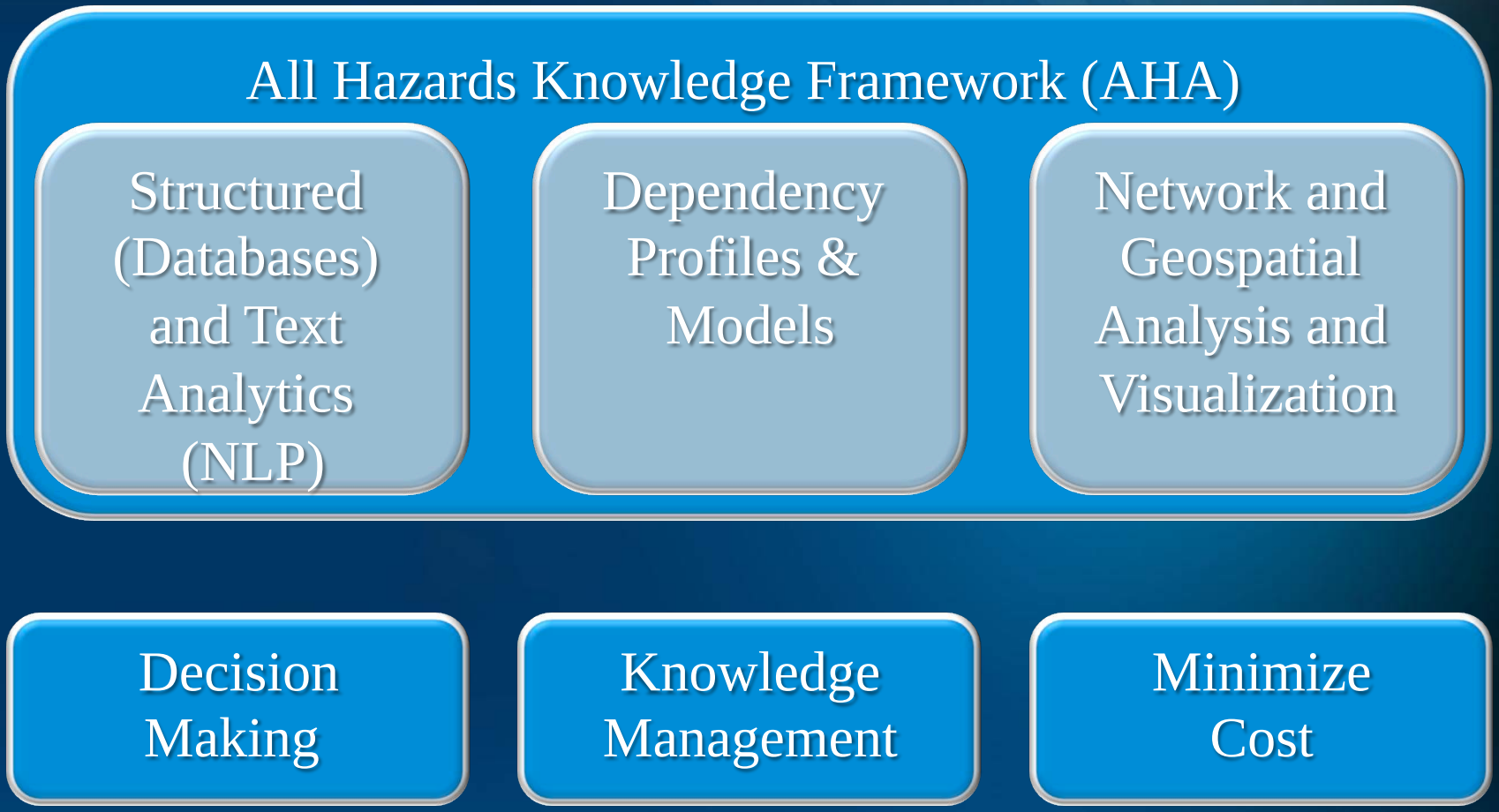
Geographic Visualization

Decision Making

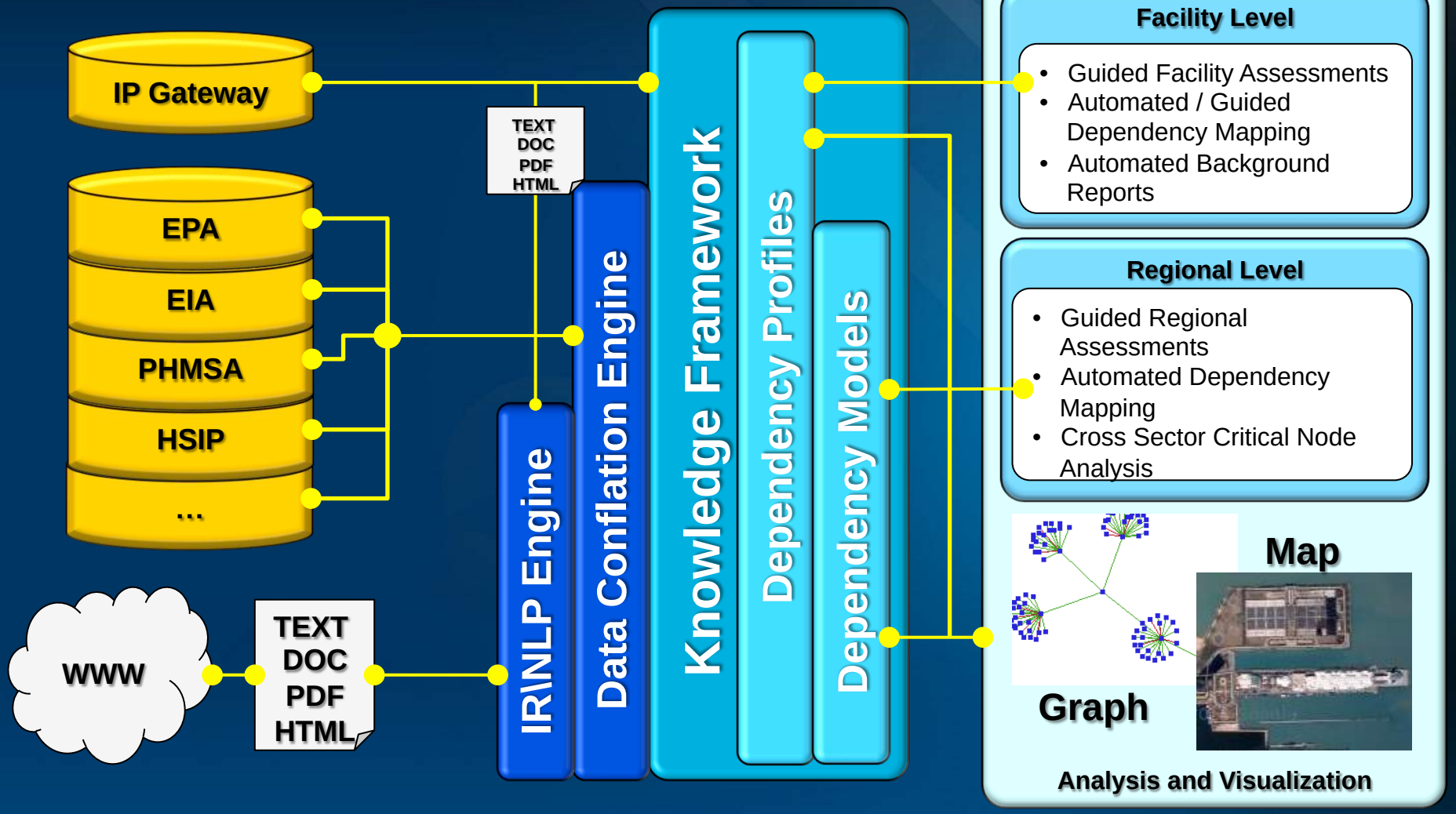
Knowledge Management

Minimize Cost

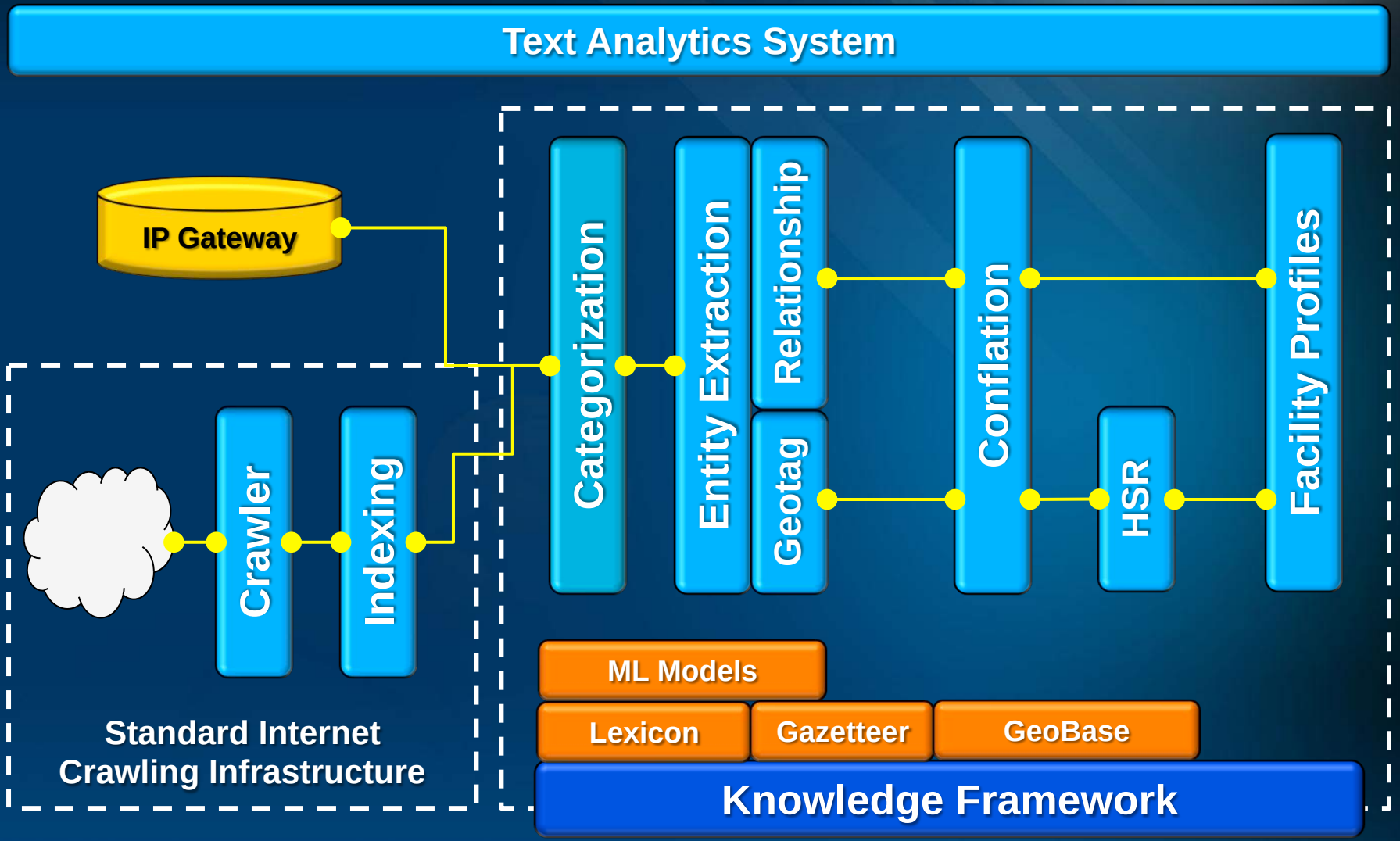
Knowledge Framework Concept



AHA Framework



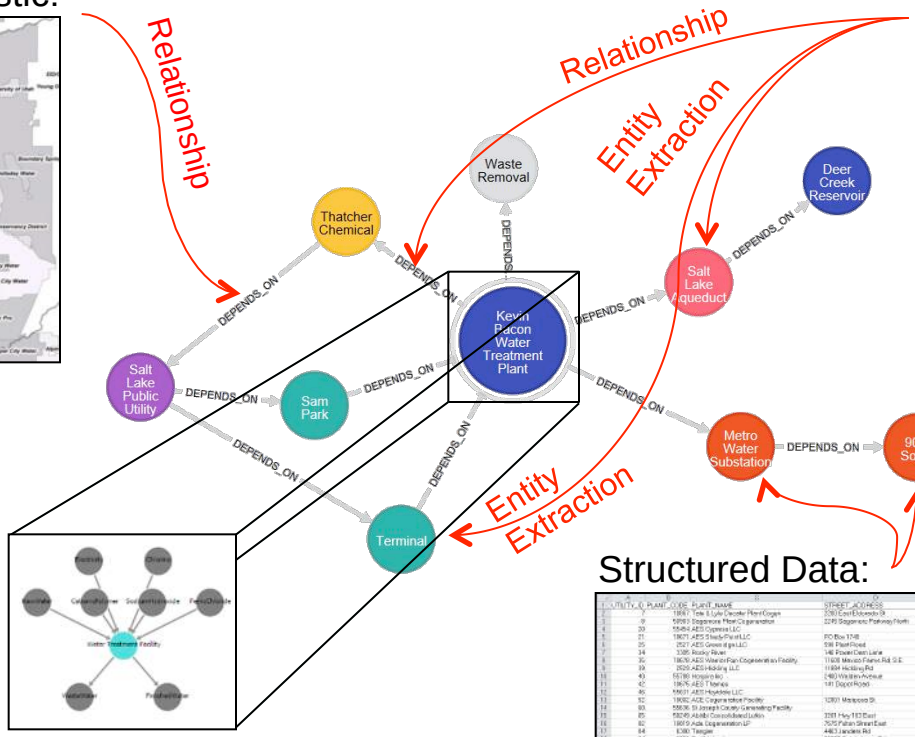
AHA Framework - Unstructured Data



Technical Approach

Spatial Heuristic:

Nearest_to
Contained_in,
etc.



Dependency Profile

Unstructured Data:

News Articles, RFPs,
After Action Reports, etc.

Attribute Assignment

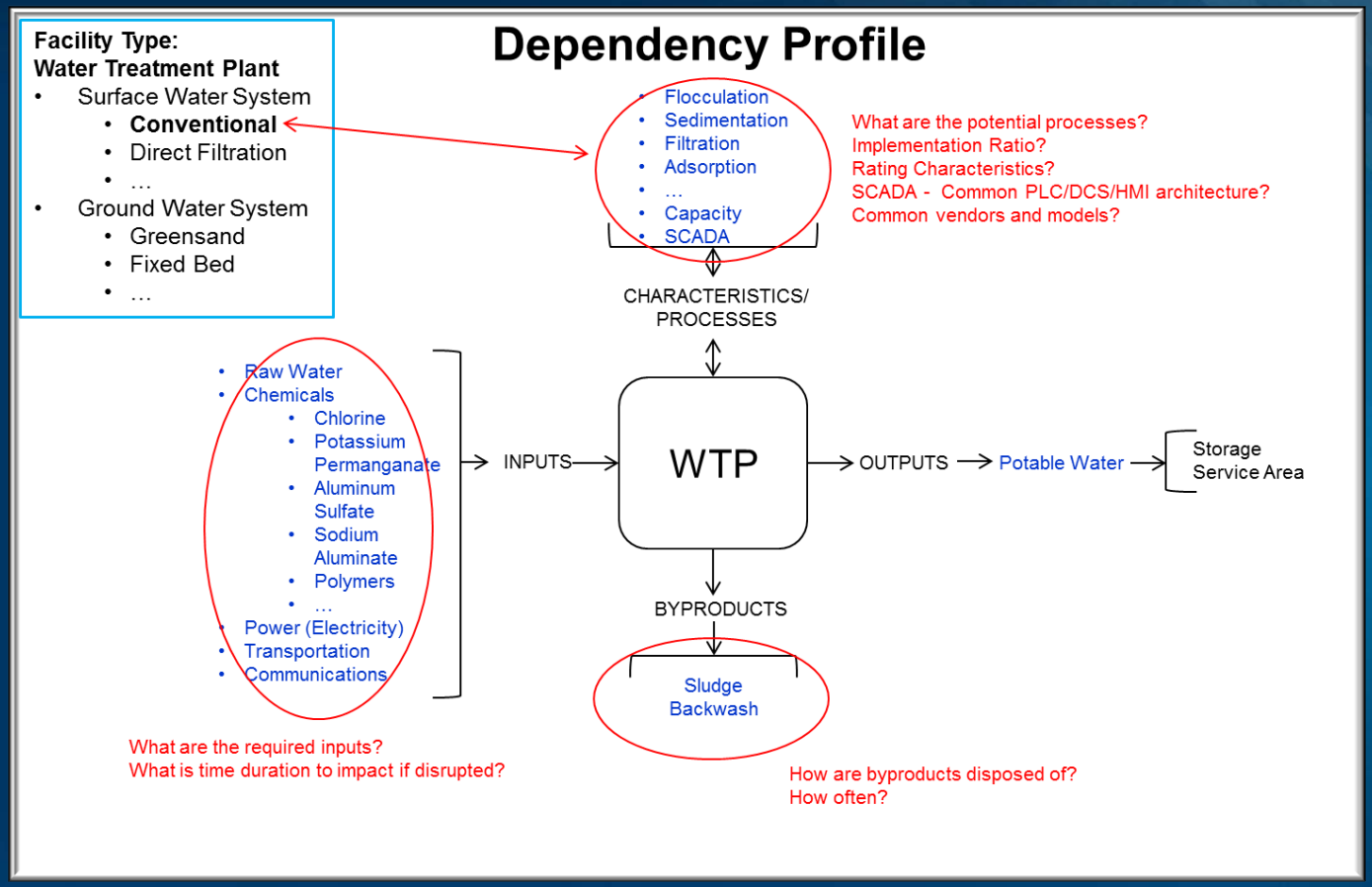
Structured Data:

Decision Making

Knowledge Management

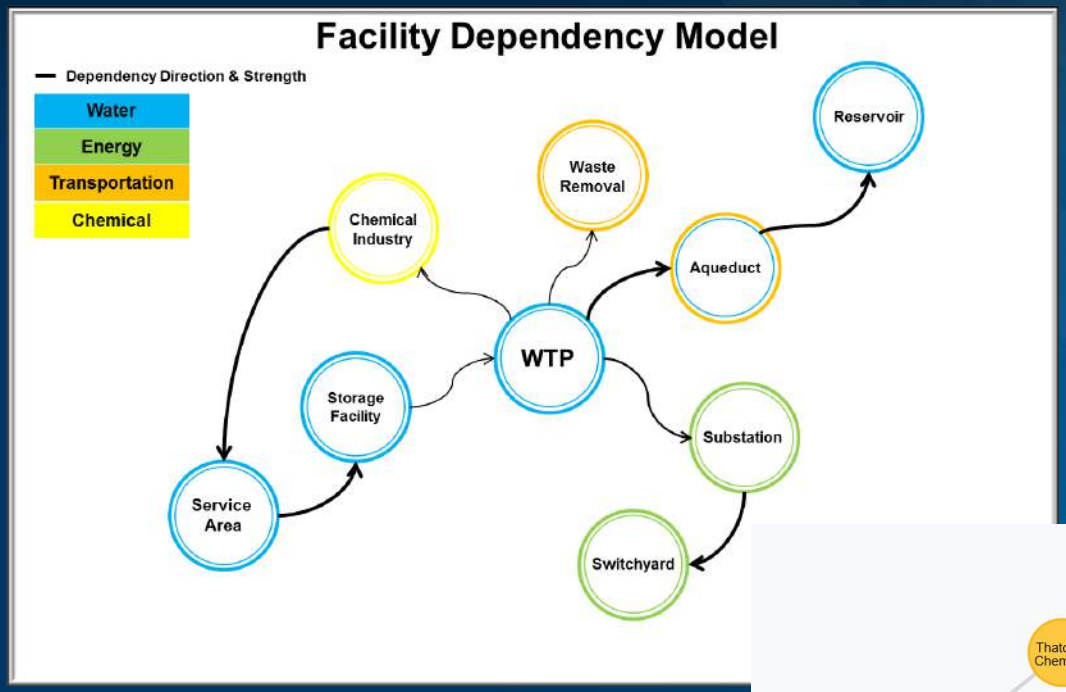
Minimize Cost

Dependency Profiles & Models

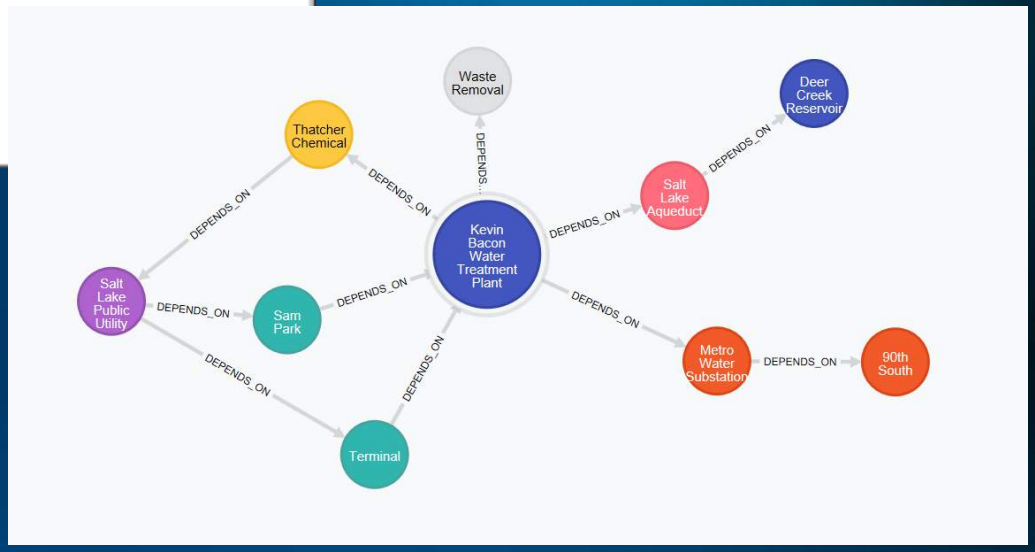


- *SME & Data Driven*
  - *I/O*
  - *Processes*
  - *Strength*
  - *Direction*
- *Profiles to date:*
  - *Water*
  - *Waste water*
  - *Electricity*
- *Lexicon (Terms)*
  - *Industry*
  - *Open Source*

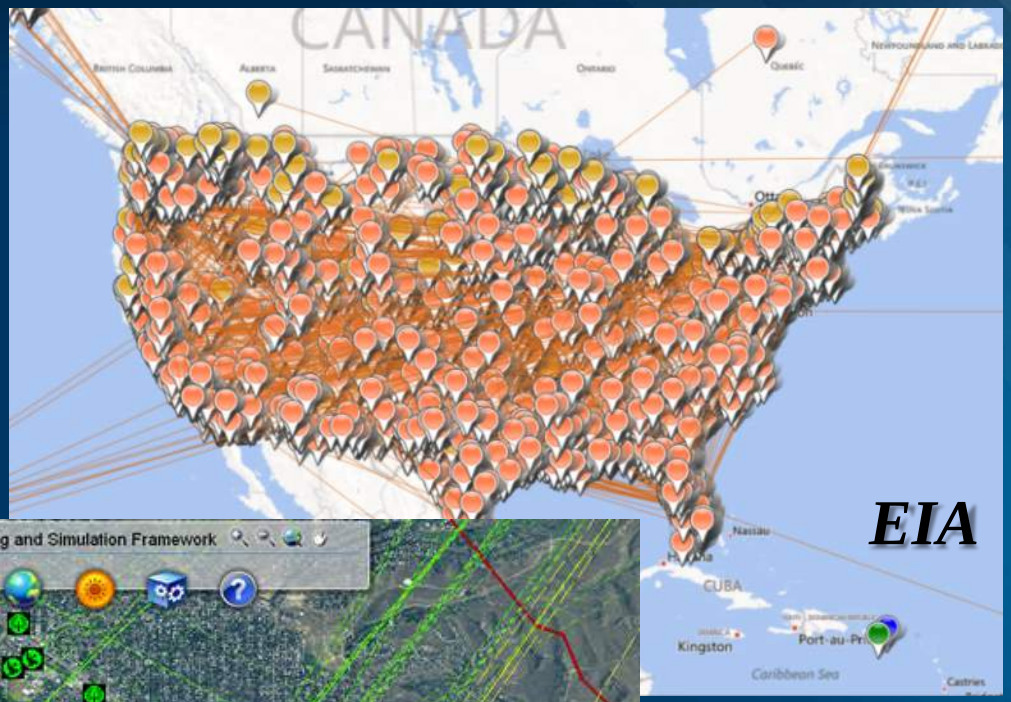
Dependency Profiles & Models

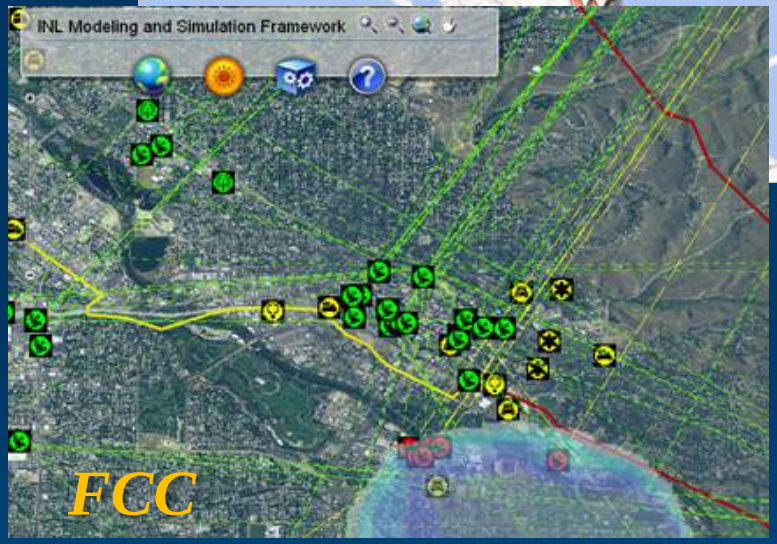


- *Survey & Assessments*
- *Validated Data*
- *Other Govt. Data*
- *Open Source*
- *Heuristics*



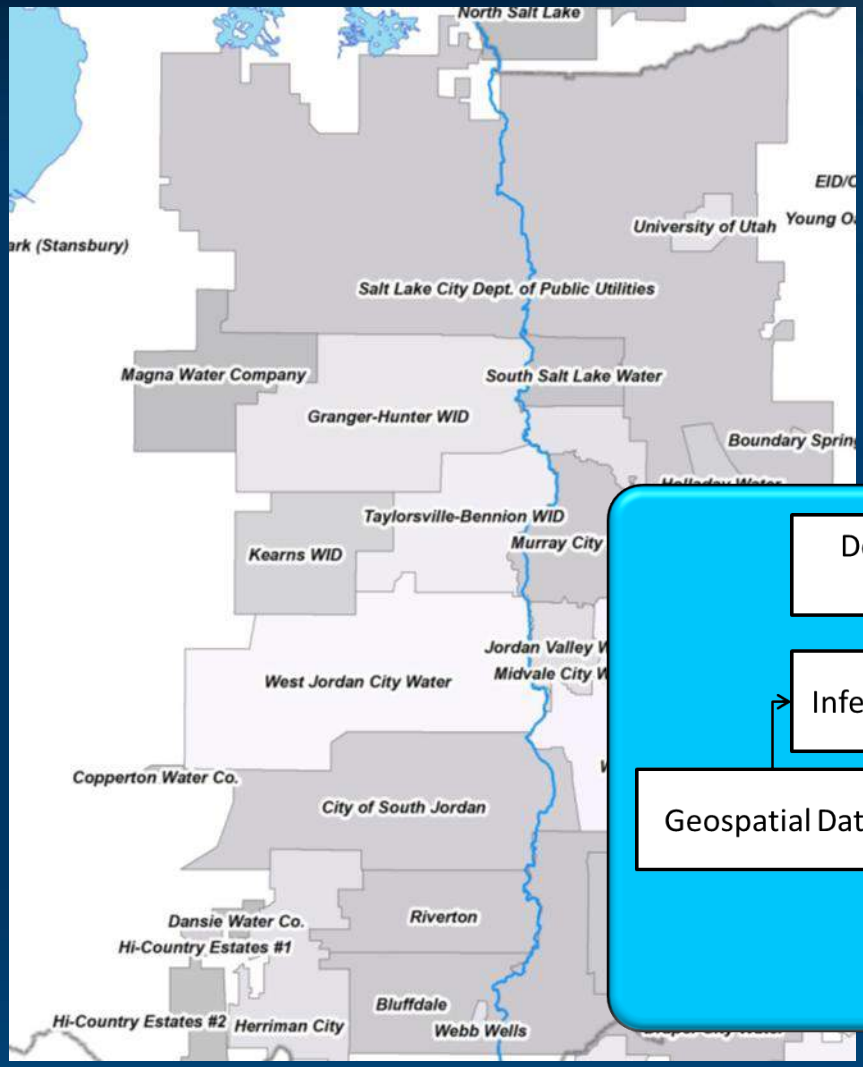
Structured Data – Other





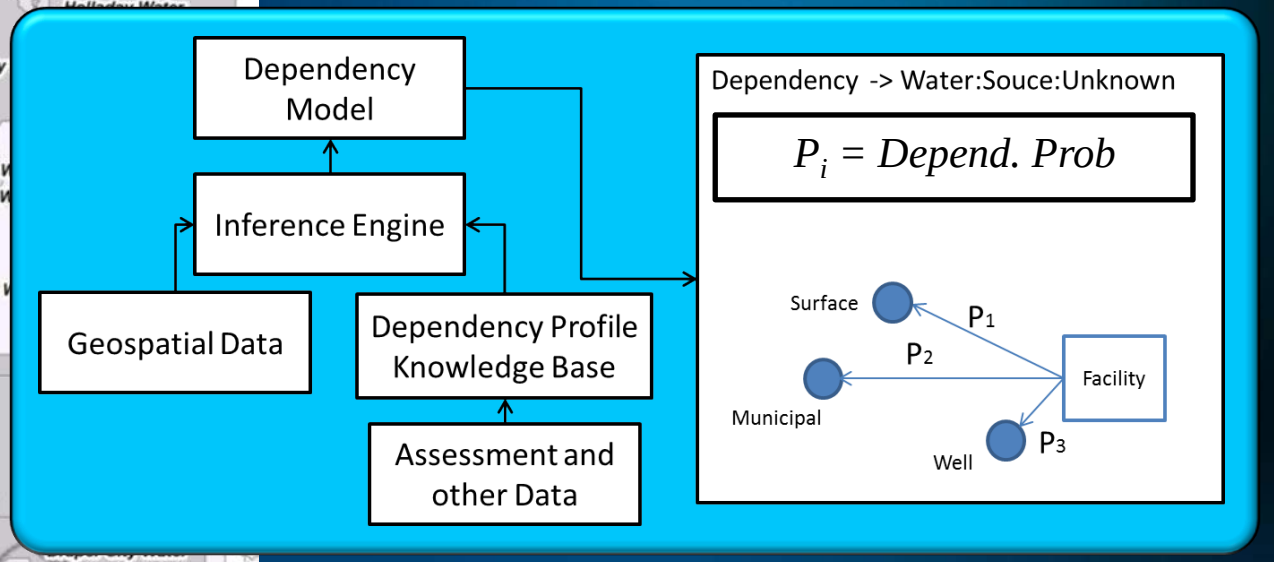
- *Straight Forward import and load*
  - *Map Fields to Profiles*
  - *Data Mining*
- *Data Sources to date:*
  - *EIA*
  - *FCC*
  - *Spatial Data (HSIP)*
- *Future*
  - *SCADA Network Visualization*
  - *GeoIP*

Structured Data - Heuristics



Dependency Probability

- *Spatially Weighted*
  - *Distance from Surface Water*
  - *Contained in service boundary*
  - *Well = f(S,M)*
  - *Facility Type*
    - *Type / Source / distance*



Unstructured Data - Text Analytics System

- Current estimate: approximately 80% of data is unstructured data
- Open source data has enormous potential to provide information
  - Dynamic Environment (Construction, etc.)
  - Aging Data
- Leverages machine learning techniques (Natural Language Processing, information extraction, information retrieval)
  - Named Entity Recognition
  - Geoparsing/Geotagging
  - Relationship Extraction
  - Document Categorization
- Critical Infrastructure Lexicons (General -> Sector Specific)
- Lifeline sector focused (including ICS infrastructure)

Unstructured Data – Web Content

- Wikipedia defines web content as the textural, visual, or aural content that is encountered as part the user experience on websites.
  - Aug 2012: Google announces that it had indexed over 30 trillion unique URLs and was processing 100 billion searches per month
  - Pew Research Center (2014): 87% of American adults access the internet
- Information contained in online news articles has the potential to enhance our knowledge
- Challenging due to the lack of (consistent) structure in web data

Text Analytics System

- Information Vector Extraction via NLP

$$IV_i = (T_i, S_i, P_i, C_i, I_i, A_i),\ i \in (1,2,\ldots m)$$

Where m is the maximum number of sentences, i represents the sentence sequence. In the ith sentence of a document,

IV_i represents the situation vector;

T_i represents the temporal information;

S_i represents the spatial information;

P_i represents the subject/object information;

C_i represents the causal/dependency information;

I_i represents the intentional information;

A_i represents the activity information;

Text Analytics System

By **Richard Rainey, NOLA.com | The Times-Picayune** on March 04, 2013 at 9:54 AM, updated March 04, 2013 at 4:10 PM

```
 <!-- Aka Story Package -->
                        <!-- gallery-preview -->
<!-- /gallery-preview -->
 <!-- Aka Secondary Package -->
                        <div class="adunit nomobilead" id="StoryAd"></div>

                      </div><!-- /storypackage -->
                    </div><!-- /#article_inset -->
                    <!-- Article -->

<div class="entry-content">
<p>As crews for the <a href="http://topics.nola.com/tag/sewerage-water%20board/index.html">Sewerage &amp; Water Board</a> continued fanning out
across New Orleans Monday morning to <a href="http://www.nola.com/politics/index.ssf/2013/03/boil-water_alert_for_east_bank.html#incart_m-rpt-
2">test the safety </a>of the city's drinking water, other investigators were busy zeroing in on the cause of a sudden drop in pressure Sunday
morning across the east bank.</p><p>The fire that broke out at the water board's aging Carrollton power plant compromised one of the steam
boilers the S&amp;WB uses to run its vast network of pump stations.</p>

<div id="asset-12371581" class="entry_widget_right">
       <a href="http://www.nola.com/politics/index.ssf/2013/03/how_river_water_is_turned_into.html" target="_blank" class="full-size-popup">
<img src="/static/common/img/blank.gif" class="lazy" data-original="http://media.nola.com/politics/photo/graphic-wtp-step3jpg-
93c245b52ee92e20.jpg" alt="graphic-wtp-step3.jpg" width="250" style="display: inline; "></a>
       <span class="photo-data">

          <span class="caption"> <a href="http://www.nola.com/politics/index.ssf/2013/03/how_river_water_is_turned_into.html" target="_blank"
class="full-size-popup">The water purification process</a></span>
          <span class="byline"></span>
       </span>
</div>


<p>The S&amp;amp;WB power plant produces an uncommon frequency of electricity known as 25-Hertz, also called 25-cycle, that drives the motors that
run about half the city's drinking water system, namely giant pumps that suck raw water from the Mississippi River and smaller pumps that send
purified water into the underground pipe network. The water purification plant itself runs on 60-cycle power purchased from Entergy.</p><p>The
power plant generates electricity by running tap water through enormous boilers, which create steam to run four turbines that produce 25-cycle
power. Natural gas purchased from Entergy is used to start boilers and turbines, which run singly or in tandem, depending on how much power the
system needs.</p><p>Once the equipment is up and running, it relies on self-generated steam and a series of natural gas feeds to stay online.
</p><p>Meanwhile, the S&amp;amp;WB has eight pumps located at three outflow stations on the South Claiborne Avenue campus to send purified drinking
water out to customers. The pumps maintain water pressure at about 65 pounds per square inch, a level that keeps potentially sickening bacteria
from infiltrating pipes; when pressure drops below 15 psi, health officials advise residents to boil water before using it to drink, cook or
bathe. </p><p>Investigators believe something went wrong with a natural gas line that powers the boilers, spokesman Robert Jackson said, but
the precise cause remains a mystery.</p><p>East Bank residents are advised to boil any water before drinking or cooking to gird against
bacteria that may have seeped into water pipes during Sunday morning's low-pressure period. The S&amp;amp;WB said it is safe to do laundry, clean
or run a dishwasher, as long as it has a sanitizing cycle. Healthy people can bathe, but it's advisable to keep their eyes and mouths shut.
Residents with compromised immune systems or open wounds should avoid the shower for now.</p><p>Algiers, which has a separate water
purification plant across the Mississippi River, was not affected.<br></p><p>S&amp;amp;WB investigators are staying in close contact with the state
Department of Health and Hospitals as they test the drinking water for contamination, Jackson said. Any test results aren't likely to be
forthcoming until this afternoon, leaving the boil water advisory in place for at least the next three to four hours.</p>

<div id="asset-12370832" class="entry_widget_large entry_widget_left">
    <span class="adv-photo-large">
```

Text Analytics System

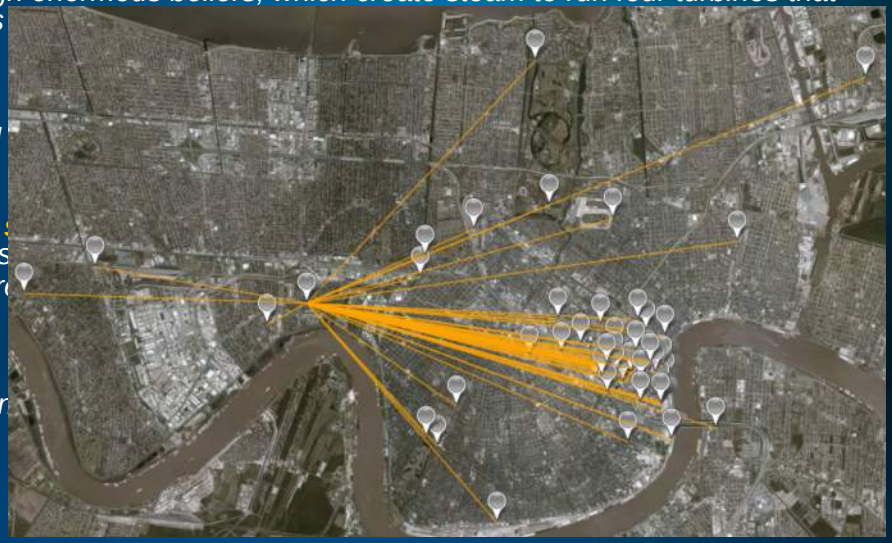
By **Richard Rainey, NOLA.com | The Times-Picayune** on March 04, 2013 at 9:54 AM, updated March 04, 2013 at 4:10 PM

*"As crews for the **Sewerage & Water Board** continued fanning out across **New Orleans** Monday morning to test the safety of the city's drinking **water**, other investigators were busy zeroing in on the cause of a sudden drop in pressure Sunday morning across the east bank.*

*The fire that broke out at the **water** board's aging **Carrollton power plant** compromised one of the steam boilers the S&WB uses to run its vast network of **pump stations**.*

*The **S&WB power plant** produces an uncommon frequency of electricity known as 25-Hertz, also called 25-cycle, that drives the motors that run about half the city's drinking **water** system, namely giant pumps that suck raw **water** from the **Mississippi River** and smaller pumps that send purified **water** into the underground pipe network. The water purification plant itself runs on 60-cycle power purchased from **Entergy**.*

*The power plant generates electricity by running tap **water** through enormous boilers, which create steam to run four turbines that produce 25-cycle power. **Natural gas** purchased from **Entergy** is depending on how much power the system needs.*



Once the equipment is up and running, it relies on self-generated

*Meanwhile, the S&WB has eight pumps located at three **outflow** drinking water out to customers. The pumps maintain **water** press potentially sickening bacteria from infiltrating pipes; when pressur before using it to drink, cook or bathe.*

*Investigators believe something went wrong with a **natural gas** lir the precise cause remains a mystery."*

…

Unstructured Data – In Action

```
Injector: starting at 2014-11-04 14:00:47
Injector: crawlDb: WaterData2/crawldb
Injector: urlDir: urls
Injector: Converting injected urls to crawl db entries.
Skipping www.slcdocs.com/utilities/NewsEvents/news1999/news051799.htm:java.net.MalformedURLException: no
Injector: overwrite: false
Injector: update: false
Injector: Total number of urls rejected by filters: 8
Injector: Total number of urls after normalization: 42
Injector: Total new urls injected: 42
Injector: finished at 2014-11-04 14:00:51, elapsed: 00:00:03
Tue Nov 4 14:00:51 MST 2014 : Iteration 1 of 2
Generating a new segment
Generator: starting at 2014-11-04 14:00:52
Generator: Selecting best-scoring urls due for fetch.
Generator: filtering: false
Generator: normalizing: true
Generator: topN: 50000
Generator: Partitioning selected urls for politeness.
Gener
Gener
Opera
Fetch
Fetch
Fetch
Fetch
Using
Fetch
Fetch
Queue
Using
```

Treatment plant power outage means all of Tampa must boil water

By Will Hobson, Jessica Vander Velde and Richard Danielson, Times staff writers
Friday, February 22, 2013 2:44pm

TAMPA — A rodent, most likely a squirrel, chewing on an electric line at Tampa's water treatment plant Friday morning caused a power failure that led to an unprecedented 48-hour citywide boil-water notice for 560,000 people and businesses and left them scrambling to snap up bottled water to last through the weekend.

RELATED NEWS/ARCHIVE

Belleair explores cheaper drinking water alternatives
3 Months Ago

In hardscrabble Homosassa, scallop shucking is a living
4 Months Ago

Hurricane season gadgets help before, and after, the storm
6 Months Ago

The effects were immediate and widespread: Water fountains disappeared under trash bags. Restaurant managers bumped into each other in lines as they stocked up on ice and bottled water. And residents across Hillsborough County tried to figure out if the notice affected them and, if it did, what exactly they could do with their tap water.

Until Monday morning, officials said, residents and businesses should boil any tap water used for drinking, cooking, washing fruits or vegetables, making ice or brushing teeth for at least a minute. This affects more than half a million people — about 100,000 of them outside city limits — who get their water from Tampa.

```
In the Learning Phase!
Indexing events using cutoff of 5

        Computing event counts...  done. 640 events
        Indexing...  done.
Sorting and merging events... done. Reduced 640 events to 635.
Done indexing.
Incorporating indexed data for training...
done.
        Number of Event Tokens: 635
            Number of Outcomes: 3
          Number of Predicates: 474
...done.
Computing model parameters ...
Performing 100 iterations.
  1:  ... loglikelihood=-703.1118647475987      0.8859375
  2:  ... loglikelihood=-318.4086912795186      0.8859375
  3:  ... loglikelihood=-234.5083746757865      0.8875
  4:  ... loglikelihood=-194.82987662624967     0.9078125
  5:  ... loglikelihood=-168.83196770130573     0.9265625
  6:  ... loglikelihood=-149.85553337665246     0.94375
  7:  ... loglikelihood=-135.27696586601488     0.95
  8:  ... loglikelihood=-123.66370298153073     0.9640625
  9:  ... loglikelihood=-114.14658808481023     0.9671875
 10:  ... loglikelihood=-106.17517699532034     0.96875
 11:  ... loglikelihood=-99.3860843543352       0.971875
 12:  ... loglikelihood=-93.52798625655316      0.971875
 13:  ... loglikelihood=-88.41907678865847      0.975
 14:  ... loglikelihood=-83.92286143636706      0.975
 15:  ... loglikelihood=-79.9340195728471       0.975
 16:  ... loglikelihood=-76.36971512506521      0.9765625
 17:  ... loglikelihood=-73.16389965592217      0.9765625
 18:  ... loglikelihood=-70.26334691224974      0.9796875
 19:  ... loglikelihood=-67.62476928728941      0.9796875
 20:  ... loglikelihood=-65.21266325388898      0.98125
 21:  ... loglikelihood=-62.99767371771704      0.984375
 22:  ... loglikelihood=-60.955339980660796     0.9875
 23:  ... loglikelihood=-59.06512719718152      0.9890625
 24:  ... loglikelihood=-57.309673385000664     0.9890625
 25:  ... loglikelihood=-55.674200151162644     0.9890625
 26:  ... loglikelihood=-54.14604839273956      0.9890625
 27:  ... loglikelihood=-52.71430989760744      0.9890625
 28:  ... loglikelihood=-51.369532953153104     0.9890625
 29:  ... loglikelihood=-50.1034854191808B      0.990625
```

Text Analytics System – Named Entity Recognition

- AHA
  - Hybrid Approach – Maximum Entropy and Continuous Random fields Classifier, Gazetteer (Keywords), and Heuristics
- Corpus (300 Articles – 632 Named Entities - Manually Annotated)
- Test Case:
  - Recall: 0.69
  - Precision: 0.77
  - F-Measure: 0.73
- Geoparsing
  - Initial testing: (100 sentences -> 1000+)
    - Naïve Bayes (70% TPR / 88.33 TNR)
    - Genetic Algorithm (82.5% TPR / 98.28 TNR) -> (91.3% TPR)

Text Analytics System – Document Categorization

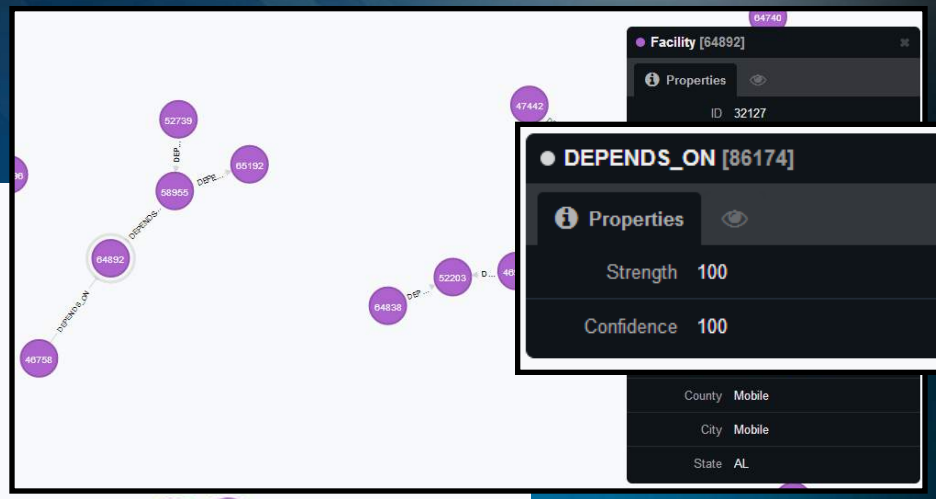
Training: 6 Occurrences										
		Predicted								
		Electricity	*IT*	*Natural Gas*	*Petroleum*	*Water*	*Wastewater*	*Other*	*Total*	*Recall*
Actual	*Electricity*	13							13	1.00
	IT		4			1		1	6	0.67
	Natural Gas		4	1	1				6	0.17
	Petroleum	1			5				6	0.83
	Water	1				24	2	3	30	0.80
	Wastewater	1				1	3		5	0.60
	Other			1				3	4	0.75
	Total Predicted	16	8	2	6	26	5	7	70	
	Precision	0.81	0.50	0.50	0.83	0.92	0.60	0.43		

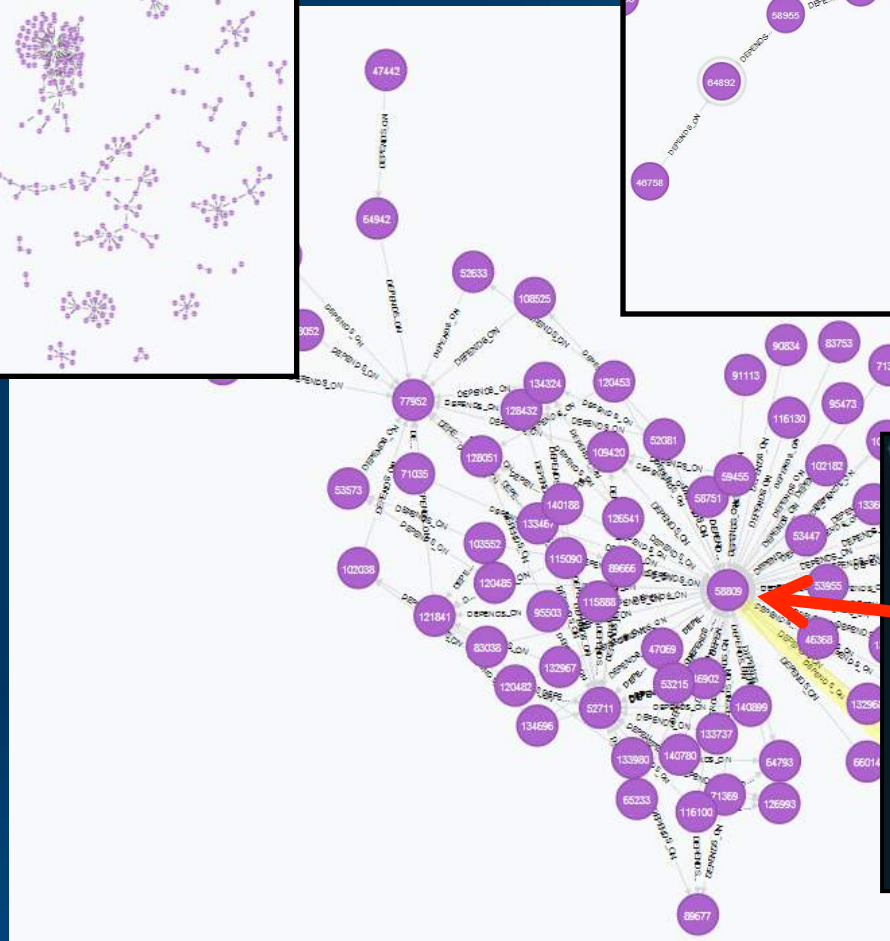
6 Occurrence Confusion Matrix

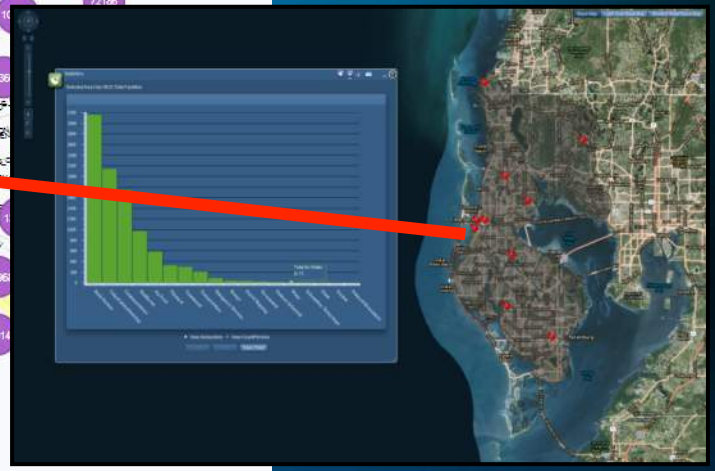
Training: 8 Occurrences										
		Predicted								
		Electricity	*IT*	*Natural Gas*	*Petroleum*	*Water*	*Wastewater*	*Other*	*Total*	*Recall*
Actual	*Electricity*	12						1	13	0.92
	IT		4			1		1	6	0.67
	Natural Gas		3	2	1				6	0.33
	Petroleum	1			5				6	0.83
	Water	1				25	2	2	30	0.83
	Wastewater	1				1	3		5	0.60
	Other			1				3	4	0.75
	Total Predicted	15	7	3	6	27	5	7	70	
	Precision	0.80	0.57	0.67	0.83	0.93	0.60	0.43		

8 Occurrence Confusion Matrix

Analysis and Visualization







Thank You

Ryan.Hruska@inl.gov