Knowledge Graphs and Disaster Mitigation

Ontology Summit 2022

We provide area briefings within seconds for any region on Earth
Can We Make Useful Data Re-usable?

Probably the most pressing foundational challenge in (meta) data management.
Data Acquisition Bottleneck

The typical data science project spends a majority of all resources on data discovery, entry, cleaning, and integration—instead of actual analysis.

80% Processing Data

20% Deriving Insights
Our Value Proposition

KnowWhereGraph aims at providing area briefings within seconds for any region on Earth to answer questions such as

- “What is here?”
- “What happened here before?”
- “Who knows more?”
- “How does it compare to other regions or previous events?”

By doing so, we assist decision-makers and data scientists in rapidly enriching their data with millions of connected, up-to-date facts at the human-environment interface to gain the situational awareness required for good decision-making.

Our ultimate goal is to flip the 80/20 bottleneck on its head.
High situational awareness is a prerequisite to good decision-making. We aim at helping decision-makers and data scientists to fast-forward through the perception and comprehension phases.
Overcoming the 80/20 Bottleneck with (Esri's) **GeoEnrichment**

Enrich (your own) data by adding (demographic) facts about your spatial area of interest.

<table>
<thead>
<tr>
<th>Year</th>
<th>State</th>
<th>Agency</th>
<th>Unit ID</th>
<th>Fire Name</th>
<th>Alarm Date</th>
<th>Cont Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>CA</td>
<td>USF</td>
<td>VNC</td>
<td>THOMAS</td>
<td>12/01/2017, 4:00 PM</td>
<td>1/11/2018, 4:00 PM</td>
</tr>
</tbody>
</table>

**Have a smartphone**

<table>
<thead>
<tr>
<th>2019 Seniors (Age 65+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8,288.00</td>
</tr>
<tr>
<td>2,480.00</td>
</tr>
</tbody>
</table>
Pros

- On-demand
- Well-curated
- Apportioned
- GIS-ready

Cons

- Pre-defined categories
- Essentially closed data silos
- Flat, tabular data
- Not always up-to-date
- Does not scale
- Limited support for automated integration
Knowledge Graphs

- **Emphasize relationships** over (in addition to) attributes
- **Connect data** (not datasets) across themes and silos
- Break apart the data-metadata distinction by making data smart, self-descriptive
- **Machine-readable**, rich semantics enables conflation and inference
- While **places** are key nodes in KG, we know almost nothing about them
Key Ideas

- Relating (sub)classes and relations (Disaster \(<\) Hazard)
- Relating entities (Vienna in graph1 is the same as Wien in graph2)
- What about the relation between Disaster in G1 and Disaster in G2?
- Gaining a more holistic understanding!
We are the first to combine geoenrichment with knowledge graphs to power environmental intelligence applications.
Pilots and Prototypes

Farm to Table Supply Chain & Sustainability
Enhance the sustainability, efficiency, and safety of consumer food supply in collaboration with the Food Industry Association.

Land Valuation and Risk of Default
Driver-based land potential assessment for model based valuation and risk assessment for agricultural credit applications & loan portfolio monitoring.

Community Lifeline System Backbone
Provide a data library for time-critical situational awareness assessment by delivering data for an affected area with over 25 data layers in seconds.

Humanitarian Aid
Apply our technologies to the humanitarian supply chain needs of Direct Relief during the COVID-19 crisis and help them to find experts.
**Hurricanes and COVID-19**

"Many Displaced from Hurricane Laura Are Now in Path of Hurricane Delta"

**Tasks**

- Find predicted path of the hurricane
- Find previous events along the path
- Understand COVID situation in x days
- Find experts for identified perils

![Without KnowWhereGraph GeoEnrichment](chart1)

**With KnowWhereGraph GeoEnrichment**

![With KnowWhereGraph GeoEnrichment](chart2)

(*) per region, e.g., state or county.
WHAT

HOW
The 30,000 Foot View

KnowWhereGraph serves 8+ different types of globally unique place and region identifiers and 10+ different thematic layers containing millions of facts about these regions.

- Discrete, Hierarchical Global Grid System
  S2 Cells (L9 globally, L11 USA, L14 CA, L16 for urban (*)
- Global Administrative Areas
- US Federal Judicial District (*)
- National Weather Zones
- FIPS Codes
- DMA (*)
- ZIP (*)
- Climate Division

Our region identifiers are linked to Wikidata/Linked Data Cloud. A team/customer that uses our identifiers gets access to billions of triples outside in addition to ~1.2B from the current KWG.
About 12-15 Billion triples (possibly spatiotemporally-scoped) extracted and interlinked across 27 data sources at the interface between humans and the environment from 12 organizations (e.g., gov. agencies, universities, and NGOs). More data is being added constantly. Final size: 30-50B
<table>
<thead>
<tr>
<th>Dataset Name/Theme</th>
<th>Source Agency</th>
<th>Key Attributes</th>
<th>Spatial Coverage</th>
<th>Temporal Coverage</th>
<th>Place-Centric Dataset</th>
<th>Defining Authority</th>
<th>Spatial Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil Properties</td>
<td>USDA</td>
<td>soil type, farmland class</td>
<td>Targeted regions in US</td>
<td>Current</td>
<td>S2 Cells</td>
<td>Google</td>
<td>Lvl 9 (Global), Lvl 13 (US),</td>
</tr>
<tr>
<td>Wildfires</td>
<td>USGS, USDA, USFS, NIFC</td>
<td>wildfire type, burn severity, num. acres burned, contained date</td>
<td>US</td>
<td>1984–current</td>
<td>Global Administrative Areas</td>
<td>University of Berkeley, Museum of Vertebrate Zoology and the International Rice Research Institute</td>
<td>Global</td>
</tr>
<tr>
<td>Earthquakes</td>
<td>USGS</td>
<td>magnitude, length, width, geometry</td>
<td>Global (mag. over 4.5)</td>
<td>2011-01-01 to 2022-01-18</td>
<td>Global Administrative Areas</td>
<td>University of California, Berkeley, Museum of Vertebrate Zoology and the International Rice Research Institute</td>
<td>Global</td>
</tr>
<tr>
<td>Expert - General</td>
<td>KWG, UC System, DR, Semantic Scholar</td>
<td>name, affiliation, expertise with spatiotemporal scopes</td>
<td>Global</td>
<td>unlimited</td>
<td>National Weather Zones</td>
<td>NOAA</td>
<td>US</td>
</tr>
<tr>
<td>Cropland Types</td>
<td>USDA</td>
<td>crop types (raster data)</td>
<td>US</td>
<td>2009-2021</td>
<td>FIPS Codes</td>
<td>NRCS</td>
<td>US</td>
</tr>
<tr>
<td>Smoke Plumes</td>
<td>NOAA</td>
<td>daily smoke plumes extent</td>
<td>US</td>
<td>2010-2021</td>
<td>ZIP</td>
<td>ZCTA</td>
<td>US</td>
</tr>
<tr>
<td>Climate Observations</td>
<td>NOAA</td>
<td>temperature, precipitation, PDSI, PHSI</td>
<td>US</td>
<td>1950 - 2020</td>
<td>Climate Division</td>
<td>NOAA</td>
<td>US</td>
</tr>
<tr>
<td>Disaster Declaration</td>
<td>FEMA</td>
<td>Designated Area, Program, amount/Approved, Program Designated Date</td>
<td>US</td>
<td>1953 - 2021</td>
<td>Census Metropolitan Area</td>
<td>US Census</td>
<td>US</td>
</tr>
<tr>
<td>Smoke Plume Extents</td>
<td>NOAA</td>
<td>Smoke extent</td>
<td>US</td>
<td>2017 - 2021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BlueSky Forecasts</td>
<td>Bluesky</td>
<td>PM10, PM5</td>
<td>US</td>
<td>2021-12-08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[This overview is outdated]
Family of satisifiability-based reasoning services

(Concept) subsumption: $KB \models C \subseteq D$

Instance checking: $KB \models C(a)$

- Requires (almost) no data
- No uncertainty at the cost of the inability to handle noise

E.g., learning embeddings as low(er)-dimensional vector (space) representations of entities and their relations

- Link prediction: predicting unknown/missing statements between (two) entities.
- Node classification: predict the class membership of unlabeled entities

- Requires a lot of (unbiased/representative) data
- Great ability to handle uncertainty and noise
Some Simple Intuitions

- contains(CA,SB) and spatiallyRelated(SB,TF) then spatiallyRelated(CA,TF)
- How do you *axiomatise* concepts such as *Democracy*? And *who* gets to write the axioms?

Classical Knowledge Representation and Reasoning

Cell

\[
\begin{align*}
\text{Cell} & \subseteq \text{Geometry} \quad (1) \\
\text{Cell} & \subseteq \text{VspatialRelations.Geoemtry} \quad (2) \\
\text{Cell} & \subseteq \geq \text{ spatialRelations.Cell} \quad (3) \\
\text{Cell} & \subseteq \geq \text{ isAdjacentTo.Cell} \quad (4) \\
\text{Cell} & \subseteq \text{ VisAdjacentTo.Cell} \quad (5) \\
\exists \text{isAdjacentTo.Cell} \land \text{Cell} & \subseteq \text{ Cell} \quad (6) \\
\text{Cell} & \subseteq \geq \text{ contains.Cell} \quad (7) \\
\text{Cell} & \subseteq \text{ Vcontains.Cell} \quad (8) \\
\exists \text{contains.Cell} & \subseteq \text{ Cell} \quad (9) \\
is\text{FullyContainedIn} & \subseteq \text{ contains}^{-} \quad (10) \\
\text{contains} \circ \text{contains} & \subseteq \text{ contains} \quad (11) \\
is\text{FullyContainedIn} \circ \text{FullyContainedIn} & \subseteq \text{isFullyContainedIn} \quad (12) \\
\text{spatialRelations} \circ \text{hasGeometry}^{-} & \subseteq \text{spatiallyRelated} \quad (20) \\
\text{spatiallyRelated} \circ \text{hasAttribute} & \subseteq \text{hasAttribute} \quad (21) \\
\text{contains} \circ \text{spatiallyRelated} & \subseteq \text{spatiallyRelated} \quad (28)
\end{align*}
\]
Learning Representations (for Types of Places)
The Issue

With Portals
The Issue
With Portals
The Difference We Make

• We don't serve links to datasets; we serve every single data record in these datasets in a searchable and AI-ready way.
• We don't provide yet another region (e.g., county) specific portal; we serve data across many different types of region identifiers.
• We don't make you download, overlay, and analyze different datasets on various topics relevant to your information needs; we answer your questions directly and based on pre-integrated data.
• We don't go offline with the next technology change, e.g., Flash, Silverlight; we rely on a distributed, internationally standardized open-source technology paradigm.
Spatiotemporally-explicit KG embeddings that are invariant under syntactic changes

Learning geographic knowledge graph summaries for apportioned geo-enrichment

Quantifying and representing regional differences in the conceptualization of geographic space

How to detect and mitigate bias in geographic knowledge graphs
Team and Partnership