Blog Post 2: Data as the Engine - Fueling PANTHEON's Digital Twin

A Smart City Digital Twin (SCDT), like the one being developed in the PANTHEON project, is fundamentally data-driven. Its ability to accurately mirror the real world, simulate potential disasters, and provide actionable insights depends entirely on the quality, structure, and analysis of the data it consumes. This post explores PANTHEON's strategy for managing diverse data sources and leveraging statistical analysis to ensure the realism and reliability of its simulations.

Handling Diverse Data Streams

The PANTHEON SCDT must integrate information from a wide array of sources relevant to disaster scenarios in our pilot regions (Attica, Greece, and Vienna, Austria) . This includes:

- **Geospatial Data:** GIS layers, topographical maps, satellite imagery (e.g., Copernicus GeoTIFF, NetCDF).
- **Sensor Data:** Real-time inputs from weather stations, seismic monitors, structural health sensors, IoT devices, and environmental monitors.
- **Infrastructure Data:** Maps and status information for utilities (power, water, gas), transportation networks, and critical facilities.
- **Statistical & Demographic Data:** Population density, vulnerable group locations, historical disaster records (often in CSV or tabular formats).
- **Operational Data:** Emergency response protocols and real-time inputs from first responders.

PANTHEON employs a flexible **data representation** strategy, utilizing relational databases (like PostgreSQL) for structured data (JSON, CSV) and object storage systems (like Minio) for geospatial files . Data presented in less structured formats, like HTML web pages, is parsed and stored appropriately. This ensures data is accessible and usable by various system components.

The Role of Statistical Analysis

To move beyond simply reflecting the present, the SCDT must simulate potential future events realistically. **Statistical analysis** of historical data is crucial for this . By examining past occurrences of hazards like wildfires, earthquakes, or heatwaves, PANTHEON identifies key statistical parameters and probability distributions that characterize these events.

For example, the *frequency* of wildfire occurrences might follow a **Poisson distribution**, while the *time between* events could be modelled using an **Exponential** or **Weibull distribution** . The *size* of wildfires often follows a **Log-Normal distribution** (many small fires, few very large ones) , and *extreme events* might be captured by **Pareto** or **Gumbel distributions** .

These statistical insights allow PANTHEON to generate synthetic but plausible scenarios for training and planning simulations, ensuring they reflect the inherent variability and patterns observed in real-world disaster data.

With robust data handling and statistically informed models, the PANTHEON SCDT is equipped to simulate complex scenarios. Our next post will explore the core conceptual models that define *how* the digital twin approaches the different phases of disaster management.