Blog Post 2: Harvesting Big Data for Disaster Resilience - PANTHEON's Sources

In our previous post, we discussed the foundational data model and architecture of the PANTHEON Smart City Digital Twin (SCDT). Now, let's explore the diverse **types of data** this system harvests and the **sources** they come from. Effectively simulating complex disaster scenarios requires integrating information from the ground, the air, and even space.

Main Data Categories in PANTHEON

The SCDT primarily integrates data falling into four broad categories:

1. **In-Situ Data:** Information collected directly from the environment using ground-based sensors.
    - **Weather Stations:** Providing real-time temperature, humidity, wind speed, precipitation, etc., typically in CSV or JSON NGSI formats. This data is vital for modelling environmental conditions during events like heatwaves or wildfires. No personal data is involved.
2. **Aerial Data (UAVs):** High-resolution imagery and sensor readings captured by Unmanned Aerial Vehicles (drones).
    - **Imagery:** JPEGs, PNGs, or TIFFs offering detailed views for infrastructure inspection, damage assessment, or monitoring disaster spread (e.g., wildfire boundaries).
    - **Sensors:** Drones can also carry weather sensors. Again, privacy is maintained as no personal data is collected.
3. **Satellite Data (Copernicus):** Earth observation data provided by the EU's Copernicus program.
    - **Imagery & Environmental Data:** Used for large-scale environmental monitoring, land cover mapping, risk assessment, and tracking widespread events like floods or large fires. Data formats vary but are processed for SCDT use. No GDPR issues arise.
4. **Third-Party Data:** Information sourced from existing external databases.
    - **Traffic Data:** Real-time or historical data on traffic flow, speed, and congestion from roadside sensors in Athens and Vienna. Crucial for modelling mobility during evacuations or response efforts. Vehicle plate numbers or other personal data are not included.
    - **Community Data:** Statistical demographic data (population features, trends) used to understand community vulnerabilities. This data is strictly statistical and anonymized, containing no personal information.

Data Usage in PANTHEON's Disaster Scenarios

The specific data types utilized depend on the disaster scenario being addressed in our pilot cities:

- **Athens Scenarios**:
    - **Earthquake (DS-ATH-A):** Primarily relies on **Aerial (UAV)** and **Satellite** data for damage assessment and response planning .
    - **Wildfire (DS-ATH-B):** Uses **In-Situ** weather data, **Aerial (UAV)** imagery/sensors, and **Satellite** data for monitoring, simulation, and training first responders.
    - Both scenarios also leverage **Third-Party** traffic and community data.
- **Vienna Scenarios**:
    - **Heatwave (DS-VIE-A):** Mainly uses **In-Situ** weather data and **Satellite** data for monitoring and early warning.
    - **Man-Made Disaster (Cyberattack/Terrorism) (DS-VIE-B):** Can utilize **In-Situ**, **Aerial**, and **Satellite** data to support training exercises for response coordination.
    - **Third-Party** data (e.g., traffic, community statistics) is also relevant here.

Data Aggregation: Getting Data into the SCDT

Data reaches the PANTHEON platform in various ways:

- In-situ sensor data often goes via cloud platforms (like FINOT) before being uploaded.
- Aerial and third-party data are typically stored locally first, then uploaded via the platform's User Interface (UI).
- Satellite data is accessed from repositories (like Zenodo) and then uploaded.

This multi-pronged harvesting approach ensures that the SCDT can access the rich variety of data needed to build comprehensive digital representations of disaster situations. Our next post will focus on what happens *after* the data arrives: the crucial **preprocessing** steps that turn raw data into analysis-ready insights. ⚙️