Blog Post 2: Provenance & Precision – The Data Ecosystem

The Challenge of Heterogeneity A digital twin for disaster management cannot rely on a single data source. It must fuse authoritative static data with chaotic real-time inputs. PANTHEON's repository manages a complex ecosystem of data types, classified by structure (structured vs. unstructured), origin (sensor vs. synthetic), and sensitivity.

Authoritative Data Sources We integrate data from top-tier institutional providers to ensure simulation validity:

- Demographics: UN/WorldPop provides 100-meter resolution population grids allowing us to map density down to the city block.
- Environment: Copernicus Sentinel-2 imagery (10m resolution) and NASA SRTM elevation models (30m resolution) provide the physical terrain.
- Local Authority: Critical inputs like Building Vulnerability Assessments and Seismic Hazard parameters come directly from KEMEA (Greek Emergency Management Agency).
- Specialized APIs: We utilize Google Solar API for radiation flux in heatwave scenarios and OpenWeatherMap for real-time atmospheric conditions.

Data Provenance and "Freshness" In emergency response, stale data is dangerous data. The repository implements strict metadata tracking:

- Static Reference: Administrative boundaries (Updated rarely/annually).
- Periodic Updates: Satellite imagery and demographic shifts (Updated monthly/yearly).
- Dynamic Streams: Traffic and Weather (Updated hourly or real-time).

Ingestion Pipelines Data doesn't just "appear" in the repository. It flows through standardized ETL (Extract, Transform, Load) pipelines. Whether it's a CSV of nursing home locations or a GeoTIFF of wildfire fuel loads from the ArcFuel project, every dataset is validated for spatial bounds (WGS84), timestamp consistency, and schema compliance before being indexed for simulation use.