

Blog Post 1: The Data Foundation of PANTHEON's Digital Twin

A Smart City Digital Twin (SCDT) designed for disaster resilience is fundamentally dependent on the quality and integration of its data. The PANTHEON project is building such a platform to optimize Disaster Risk Management (DRM) by simulating scenarios and supporting better decision-making. For the SCDT, data is the essential fuel that allows it to accurately represent the physical environment and model complex events.

This first post in our series explores the crucial role of data within the PANTHEON SCDT, outlining the overall system architecture and the goals behind our integrated data model.

Why Data is Central to the SCDT

The PANTHEON SCDT's effectiveness depends on its ability to aggregate and process large volumes of diverse data, often referred to as **big data**. This data allows the SCDT to accurately represent the physical environment, simulate complex phenomena like wildfires or earthquakes, and ultimately serve as an effective training and preparedness tool. Handling this data requires a carefully designed **integrated data model** that defines how information flows from various sources into the system.

PANTHEON's SCDT Architecture: An Overview

As detailed in our project's architecture documents (D3.7), the SCDT gathers data from multiple sources through specific **aggregators**. This data is then sent to a central **preprocessing component** within the SCDT. This initial component cleans, formats, and prepares the data before it's used by the simulation engines and decision support tools deeper within the system. The process uses a **REST API** for data upload, making the system flexible regarding data sources and file sizes.

Data Integration Goals

Our primary goal for data integration is efficient **management and storage** of the various data types (like sensor readings, images, traffic counts) within a unified system. We use the **MinIO object storage system** to keep data organized, accessible, and secure. Key objectives include:

- **Standardized Storage:** Ensuring diverse formats (CSV, JSON, images) are stored consistently.
- **Metadata Management:** Cataloging data with clear metadata (origin, type, timestamp, location) for easy searching and retrieval.
- **Data Quality:** Implementing validation checks during ingestion to ensure reliability.
- **Unified Access:** Allowing different system components to access and correlate data from multiple sources simultaneously.

To achieve these goals, our technical setup relies on:

- **MinIO:** For scalable, high-availability object storage compatible with S3 APIs.
- **APIs and Web Interfaces:** For both automated and manual data access.
- **Standardized Formats:** Primarily CSV, JSON, and common image formats (TIFF, JPEG, PNG).
- **Security Protocols:** Ensuring data integrity and GDPR compliance (no personal data is processed).

This robust data foundation is essential for the SCDT's success. In our next post, we'll dive deeper into the specific types of data PANTHEON harvests and how they relate to our disaster scenarios. Stay tuned! 