



1st VMAP User Meeting 2024

SCALABLE PRINCIPLES FOR INTEGRATION PIPELINES OF SENSOR DATA AND SIMULATIONS USING KNOWLEDGE GRAPHS

Morten Meyer

SICK AG – Think Tank, Waldkirch, Germany

Providing FAIR data internally and externally is currently an important challenge for industry. VMAP, as a data standard for the transfer of data between engineering simulation tools, addresses interoperability and for the simulation domain, not unlike some standards do for sensors. We consider VMAP as part of a broader landscape: data is context-dependently required in different forms – but identity and meaning of the data itself does not change! At the example of one scalable VMAP processing architecture, we provide an overview that highlights important principles for integration of distributed data in a FAIR open world approach:

- Distributed data architecture: a data mesh with a knowledge graph that orchestrates different "perspectives" is powerful. For scaling, the enterprise's architecture must account for the fact, that knowledge domains have their own life cycle.
- Explicit meaning is crucial: at SICK, we introduced a core ontology - the SICK Application Model (SAM) for this purpose. VMAP represents what we call a 'Profile', i.e., a contract between interfacing functionalities that defines the meaning and structure of their provided or consumed data.
- Separation of meta-levels of identity and representation of data sources: A VMAP file itself is embedded in a broader application context and embodies some specific content. We can address questions about infrastructure as well as questions concerning the domain that depend on the meaning of the data object, e.g., "where can I find the temperature physically", vs. "what is temperature", both address semantic representations.
- There is only one source of truth. This reflects explicit responsibilities for data. Source data bases are sources of truth and data can only be transformed in formal processes since there must not be a second component with the same rights. One can, of course, transfer aspects the responsibility under given circumstances.

- Data pipelines like ETL are nothing but one exposure option among many. To have a second kind of exposure in this example, we can use a virtual knowledge graph for VMAP.
- Storage is not the same as access. The physical storage location and low-level management of data objects is a secondary feat for the integrator, as long as the data pipelines are built on FAIR principles and the semantics is explicitly defined. It is a primary concern only for those systems that drive the efficiency of CRUD operations at scale. Time series data, for instance, will always be stored in a different native data base than document-based data.
- Components that serve or consume data are also data objects and should be FAIR. This means that, as far as possible, micro-service-based architectures should be used for the tooling landscape. This suggests the benefit of a gateway that can expose data as VMAP for different native management systems. Following this logic, we outline a possible reference architecture for VMAP document stores for future community discussions.