Noname manuscript No.
(will be inserted by the editor)

Semantic Data Management for Experimental Manufacturing Technologies

Nasr Kasrin · **Maliha Qureshi** · **Simon Steuer** · **Daniela Nicklas**

Received: date / Accepted: date

Abstract

Experimental manufacturing technologies can significantly improve production processes. One example is the microwave manufacturing of composites, which can save energy and reduce turn-around times compared to the traditional heating in ovens. However, since this technology is not well-understood yet, it requires more research and development activities (like simulation or heating experiments) to enable stable and efficient production with controlled product quality. These activities will span multiple divisions in companies and will be carried out in an iterative manner. In this paper, we show how semantic data management can help in such experimental manufacturing technologies to bridge the gap between different activities and to carry the knowledge gained in earlier steps to later activities in the process. The approach we propose semantically enriches and catalogs both data sets and documents via a metadata triple store, alongside storing and servicing heterogeneous data (structured and unstructured) using a polyglot persistence model. Our proposed system is modular and service-oriented. It easily integrates with existing systems in the manufacturing environment via HTTP interfaces. The example use case for the architecture are simulations and experiments on microwave heating of composites as carried out by the Horizon 2020 project "SIMUTOOL"[1].

Keywords semantic, data management, knowledge management, data lakes, digital manufacturing

Contact: Nasr Kasrin
University of Bamberg
E-mail: nasr.kasrin@uni-bamberg.de

[1] http://www.simutool.com

1 Introduction

Efficient development of new and experimental technologies is a key driver of innovation and growth in production. With the advent of digital manufacturing and the so-called Internet of Things (IoT), every activity or thing has data relating to it. An opportunity here is to leverage all this data to support the evolving, iterative and collaborative processes of developing little-understood technologies. One example of such an experimental manufacturing technology is the microwave manufacturing of composites. Up until recently, traditional heating ovens have been used to cure composite parts, this technology is naturally well understood by now but has its drawbacks, the main one being energy consumption. Using microwave oven technology to cure composites is a recent development, one of its main strengths is saving up to 50% of energy relative to the traditional method. It is, however, poorly understood and there is a large gap in the know-how of curing composite parts (ex. how the micro waves interact with different materials and how the material themselves interact with each other).

To be able to fully leverage the economic benefits of this technology, an iterative process is needed that starts with the simulations of curing of the parts and products under different conditions in a microwave oven, followed by sensor-observed experiments. Finally, when the parts are under production, the process is still monitored by sensors. In each of these activities, information from previous activities (i.e., simulation results or sensor data from experiments) would be extremely helpful.

Within the scope of this project, we address three main challenges. First, how can we link documents or data from different activities so that, e.g., related simulation results can be retrieved visualized during an experiment? Secondly, how can we manage documents (written by humans, e.g., specification documents or interpretation of experiments) and data sets (produced by sensors or simulation programs) in an efficient and scalable way? And thirdly, how can we compare time series of different data sets (e.g., from simulations or previous experiments) online with an ongoing experiment or production step so that we can react quickly to unwanted situations?

The approach we propose semantically enriches and catalogs all this data via a metadata triple store, alongside storing and servicing the *Data Artifacts*[2] (data and documents) using a polyglot persistence model. Our proposed system is modular and service-oriented, and easily integrates with existing systems in the manufacturing environment via HTTP interfaces. The work presented here which was is carried out in the scope of the Horizon 2020 SIMUTOOL project.

The remainder of the paper is structured as follows: Chapter 2 discusses related work, Chapter 3 covers the processes involved in the development of experimental manufacturing technologies, Chapter 4 presents the domain model, Chapter 5 discusses the implementation of our system, its components and structure, Chapter 6 presents the online monitoring system, an application that plugs into our system and makes use of its services to produce added value, and finally Chapter 7 wraps up with some concluding remarks.

2 Related work

The work presented here shares similarities to several overlapping fields, which we will cover here. This section is divided as follows: Knowledge Management Systems will be discussed in Section 2.1, Scientific Data Management in Section 2.2, Semantic Data Management in Section 2.3, and Big Data Management in the scope of industrial environments in Section 2.4.

2.1 Knowledge Management Systems

Knowledge management (KM) has become a strategic mandate for most world class organizations. A key enabler for implementing an effective KM system is advanced information technology [16]. In particular the life-cycle of KM is about knowledge capturing, knowledge development, knowledge sharing and knowledge utilization.

[2] We use the term *Data Artifacts* to denote the both data and documents, anything store-able. An alternative term could have been digital artifacts.

2.1.1 KM in Organizations

Many organizations employ KM systems in one way or another. SIGNA International is using a group support system through 55 global units for sharing Knowledge and organizational learning across countries [3].

Knowledge is a core competency in consulting; Consulting firms are well-known pioneers in implementing KM system to support their Consultants [19]. Skandia, the large Swedish financial services company, internally audits its intellectual capital every year for inclusion in its annual report to stockholders. One goal is to persuade investors of the value of Skandia's knowledge capital. Another goal is to focus on how the organization can increase or decrease its effective use of knowledge assets over time [10].

2.1.2 KM in Theory

Alavi, et al. [2] categorizes the types of Knowledge Management Systems based on different perspectives of Knowledge that organizations have. The perspectives they present include:

1. Knowledge vis-a-vis data and information: data as facts, raw numbers, with information required to be processed and interpreted,
2. State of mind: knowledge as a mind state of the employee,
3. Object: knowledge is to be stored and managed as an object,
4. Process: knowledge as a process of implementing expertise,
5. Access to information: knowledge as the enabling of effective access to information,
6. Capability: knowledge as a capacity to affect decision making.

From this view, our work focuses on knowledge as an aggregate of (1) knowledge vis-a-vis data/information, (2) knowledge as an object, and (3) knowledge as a capability. Hence we view the process of knowledge management as the management of data and information, as the coupling of relevant knowledge as an object (as an activity in our domain model see Chapter 4), and to facilitating of decision-making.

The work in KM theory is useful in laying groundwork in defining what knowledge can mean as well as what knowledge management systems can achieve. But it falls short of giving a road-map or technological recommendations. Next we will look at another field that falls more on the applied side.

2.2 Scientific Data Management

Scientific data management as a research problem is still in its infancy, there is no clear consensus on which problem domain it targets. However, picking one dominant pattern in these systems: they are concerned with supporting the *life-cycle* of data-intensive, empirical scientific research [12]. This life-cycle is multi-staged and distributed [6] across different expertise and locations. with focus on provenance as well the 'lifeline' of datasets [4] as they are annotated, prepared, transformed, processed, etc. Example of such systems include GenBank[3] (NIH genetic sequence database), the BEXIS[4] project from the University of Jena, and from a commercial side Informatics's SDMS[5].

Scientific Data Management systems have yet to reach a tried-and-tested formula. Projects seem to work well for the domain they are targeted at (life sciences for example). But moving away from the particular empirical science environment to one like experimental manufacturing, it becomes less applicable to support such heterogeneous and multi-faceted environments.

Next we will look at systems that try to go one level higher, and deal with the semantics of data models and data sets to achieve more generality.

2.3 Semantic Data Management

Semantic Data Management is a term applied to any research that uses semantic modeling and reasoning to solve data-related challenges in the context of a data-intensive system. In the experimental manufacturing domain, the goal of Semantic Data Management is to achieve semantic integration between data coming from experiments and relational data which indexes the contents of the Knowledge Management system.

Alejandro Rodriguez et al [23] introduces the approach of semantic indexing strategy to construct RDF statements as it minimizes the impact on current Resource Description Framework (RDF) storage and query engines, and directly tailors the streaming application's actual data which is stored outside the scope of RDF. This change makes the process of data changing and updating rapid, while metadata annotation might be a little slower to update. This approach allows seamless integration of both types of data into a single indexing system.

There is has been work ontological modeling and reasoning to achieve data integration and interoperability as well as ontology-based data access [22,7,9,15]. Evgeny Kharlamov et al [14] proposed to use the Ontology Based Data Access (OBDA) approach for semantic access to streaming and static data at Siemens to enhance the direct data access. The idea behind this is to use ontologies to enrich the data model semantically, and create a mediate between user and data. Ontologies make the domain abstract and queries are used to formulate the information needed from the data, and it can be done without a domain expert. This approach hides all the lower-level technical details of data processing and leaves the user to focus on what they want to do with the data.

These systems achieve better domain independence than the Scientific Data Management Systems, however popular work around OBDA for example is focused on a specific data management problem: mainly data integration. Whereas we are looking for a semantic solution on a higher level than that of internal data structure.

2.4 Data Management in the Industry

In recent years, a variety of IoT platforms have emerged to facilitate the development and operation of complex IoT systems [5]. Many companies distribute their own IoT platforms [6,7] for specific use cases. A common use case of this IoT platforms is to capture, store, analysis and visualize data, so called Big Data. Hadoop is well known as the master solution for Big Data and a lot of platform includes it. SAP HANA Vora, for example, is developed to build a basic layer for upcoming Big Data applications in the enterprise [26]. It is becoming clear that the term Big Data is no longer equal to Hadoop [28], here for example the author describes a streaming architecture from a business value perspective with a polyglot persistence instead of a single NoSQL database.

Such Big Data, IoT industrial project do capture and facilitate the generation of added business value and control, but are harder to integrate into organization-wide knowledge management systems. Another issue, is that these Big-data solutions are targeted at data analysts, but the goal of KM systems is to be usable by many employees of an organization, regardless of their technical expertise in data science.

Whereas each of the above discussed fields achieves their goal within their domain, they are less suited to support the knowledge management of experimental manufacturing technologies. We use insights from

[3] https://www.ncbi.nlm.nih.gov/genbank/

[4] http://bexis2.uni-jena.de/

[5] https://www.coreinformatics.com/products/core-sdms/

[6] https://www.bosch-si.com/iot-platform/bosch-iot-suite/homepage-bosch-iot-suite.html

[7] https://cloud.google.com/solutions/iot/

scientific data management for understanding datasets and provenance issues, we adopt a semantic data management approach however on a more abstract level of data , documents, systems and users, and relations between them. Finally the data-intensive requirements of our domain leave us to tackle some similar problems to pure Big Data management albeit embedded within a larger system, that works transparently from the perspective of the the targeted user from the production and manufacturing side.

3 Manufacturing Technologies Development

An activity (denoted by ovals in Figure 1) is the basic unit of abstraction used in the SIMUTOOL Knowledge Management System (KBMS) to model the process of developing experimental manufacturing technologies. Example activities include: Product Design, Simulation, Experiments. Related activities can be associated by adding a relation between them. In addition to the core activities, there are several user roles in the project. Every user of the system interacts with the KBMS system under one or more roles.

A role can be understood as a hat that the user wears in regards to which activity when interacting with the system. Some users can have multiple roles (ex. Production Engineer to run experiments and take part in the production). Different roles also have different types of authorization. Product developers can create new Part instances, and Product Design activities, but cannot add or modify simulation data runs.

Figure 1 presents an overview of the activities in the Microwave Manufacturing R&D process. Below we will expound on the three activities we are most concerned with with respect to the KBMS: simulation, experiments, and production.

3.1 Simulation

The simulation activity proceeds by getting the specification of the part produced by Product Design. This includes the geometry of the part, the material specifications and additional requirements. The simulation setup will require the building of the models into CAD / CAM / CATIA and the set-up of the simulation parameters and material properties. If a new oven is being used or the oven has been upgraded then that also needs to be adjusted in the oven simulation system. The simulations are then run and runs that are deemed valuable are uploaded to the KBMS with the basic metadata properties[8].

– *Role*: Simulation operator can add or modify simulation data. Select superseding runs and delete simulation runs. They also have access to the CAD/CAM models developed by the Product Designers or Tooling Developers. The interface between the Simulation systems and the KBMS can be automated or manual. The role of simulation operator can upload new simulations, delete redundant or outdated runs.
– *Knowledge Exchange*: Simulation Operators can ask for a part, material and oven specification from the KBMS and run simulations. The results of simulation runs will be send to the KBMS with the details of the curing cycle and different variants. This information can be used for the future simulations or during the Experiment activity. A user can query the part specifications, the curing cycle and the oven specification from the previous simulation runs.

3.2 Experiments

The Experiments activity is carried out to understand the production process of the part further and to derive important parameters for production. The experiments make use of the simulation data in the monitoring of the process (more on this later). The curing cycle is one of the major inputs for the experiments. As the experiments are concluded a new Measurement activity with the definition of the curing cycle will be recorded. It is also possible to create version and supersede versions. The Experiments interface with the KBMS system via the Monitoring system (in Experiment mode) which visualizes the sensor and simulation data to assist the experiment operator in knowing the detailed status of the experiment.

– *Role*: Production Engineers can be part of several activities in the SIMUTOOL project: Measurements and Experiments. Production engineers created new experiments, adds material properties, query for simulation runs, curing cycles, and part specification. Production Engineers along with line engineers are the only roles authorized to add an evaluation to the KBMS regarding the final product. Additionally they can add and modify Oven specifications. The monitoring system can be accessed in expert mode by Production Engineers, more on the monitoring system below.

[8] We use the core element set of Dublin Core (`http://dublincore.org/`), with properties such as creator, date, descriptions. We add more elements in later parts. See Chapter 5 for more information.

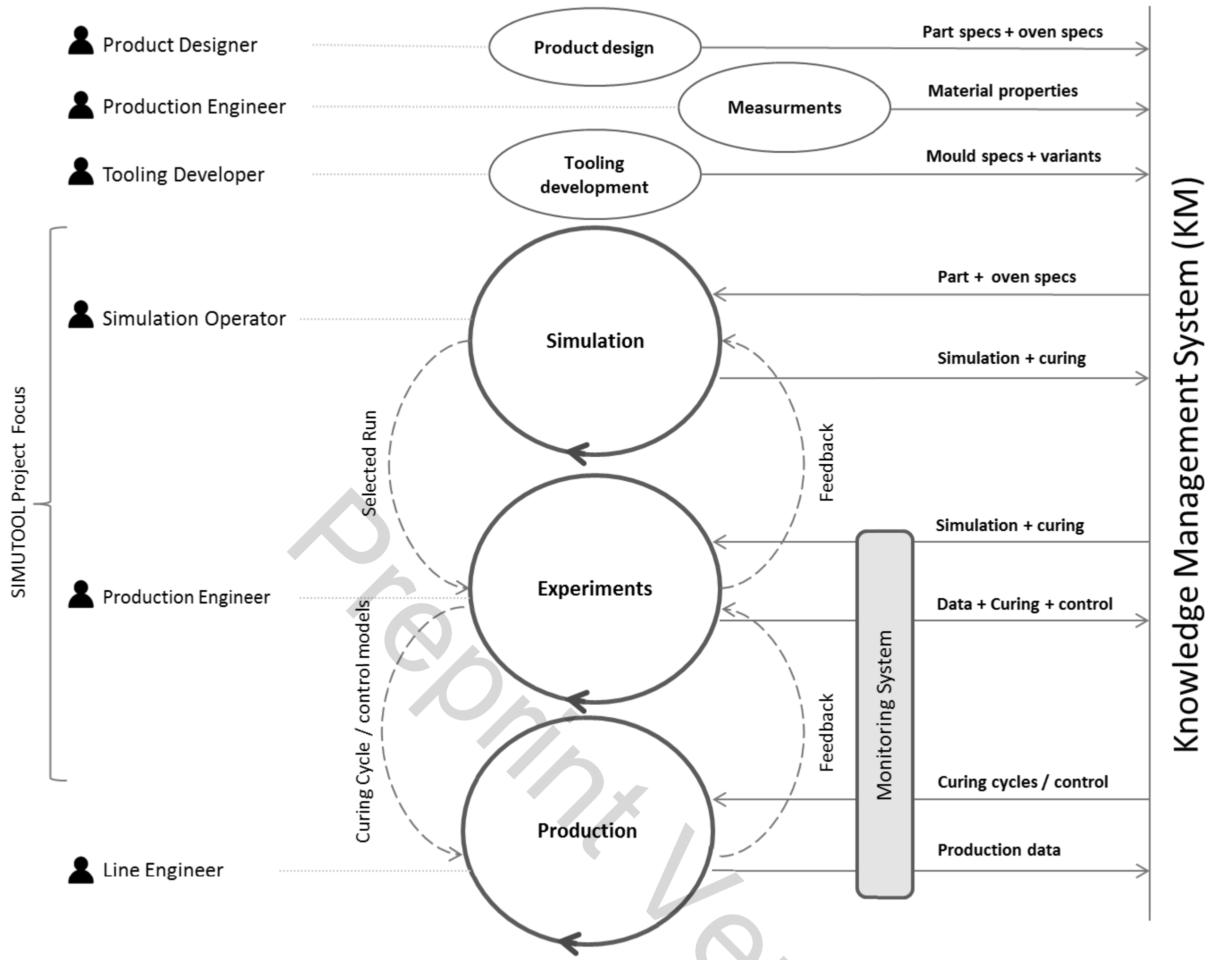

Fig. 1 Depiction of the Experimental Manufacturing Process

– *Knowledge Exchange*: Production Engineers can ask for simulation runs, curing cycle with variations of different temperatures and sensors from the KBMS. In the experiments, Production Engineers will compare the simulation runs with the live sensor data and can return the production sensor data, control models, the curing cycle and the final evaluation from the experiments to the KBMS.

3.3 Production

The goal is to control and monitor the production. The interaction with the KBMS system and this stage will also be via the Monitoring system (in production mode). The final evaluation and the sensor data of the process at the end are both uploaded to the KBMS system by creating a new Production activity instance and attaching to it the sensor data and the Part evaluation.

– *Role*: The role of line engineers affords the ability to search for simulation runs and curing cycles and to access the Monitoring system in Production Mode. The Line Engineer role also has the authorization to add final part evaluations to the KBMS at the end of production runs. The monitoring system can be accessed in production mode by Line Engineers, more on the monitoring system below.
– *Knowledge Exchange*: The line engineer will ask for the curing cycle and the control models from the experimental specifications in the KBMS. The line engineer can use it to control the production run with the experimental specification, generate the new production data for the generated parts and upload them to the KBMS for future reference.

Following from this analysis, we need a formal method to represent these ideas and concepts for the knowledge management system as well as for the end-user and domain experts. Domain modeling achieves these goals and offers us a well understood set of tools and methods to represent such information. It also helps us model the information in such a way that a semantic reasoning system can handle and manage, as well as serving as a communication platform between the KBMS system developers and the end-users / domain experts. In the next chapter we give an outline of our domain model.

4 Domain Model

The SIMUTOOL base domain model is created to define essential concepts of the knowledge base management system. The domain model defines logical structure of the domain knowledge[17].

The SIMUTOOL knowledge management domain model consists of two parts: the upper-level ontology and the lower-level ontology.

The upper-level ontology consists of the super class SimutoolThing and its subclasses: User, Document, Activity, and Resources.

- Class User consists of the information about the users involved in the different tasks in the domain. For instance a production engineer is responsible for several parts like creating new experiments, specifying material properties.
- Class Activity represents the activities in the process that can be linked to the knowledge in the documents or data set that are linked to the instances of the activities like simulation activity, measurement of a component or material activity, controlling the temperature activity, recording the experimental specification activity.
- Class Resource represents the information about the resources linked to activities and contained in documents like oven used to heat the materials, the material used in an experiment, tools used in an experiment, parts molded during an experiment.
- Class Document holds the knowledge for Data uploaded or created by different experiments like Documents related to activities.

The SIMUTOOL Knowledge Management domain is generic on the upper level that can be used to advantage for modeling business concepts like User, Activity, Documents, and Resources and can be modified according to specifications of different domains. Lower level details of domain model like Activity Instances, Resources Instances are specific to the SIMUTOOL knowledge management domain, which contains the specific information about manufacturing processes. This information can be changed according to the specifications of the end-users.

Classes and instances in the domain model communicate through object properties and data properties. Object properties represent the relationships between instances like "a document is uploaded by a user". Data properties represents the relationship between instances like a specific user and data values like email address.

5 Implementation

In this chapter we will discusses the implementation of the KBMS system; its components and structure. A depiction of the implementation of the KBMS can be found in Figure 3. We will start with an overview of the systems and go on to discuss the different parts in more detail.

5.1 Overview

- Clients: Human users, systems (ex. Simulation systems) and sensors can all use the KM system.
- Online Monitoring: A component to monitor live sensor data and compare it with simulation data.
- Push/Pull web interface: Web pages for human users to add, search and retrieve data and information.
- Knowledge coordinator: As the name implies this component is responsible to decide which part of the data and information to forward to which services
- Extractors: System the extract data from files and documents
- Semantic layering: see next section for details
- Data services: services responsible for storing and retrieving raw data and documents. This section of the system can be considered more of a federation of data management systems: relational DBMS, big data stores, NoSQL, file systems, etc. In the cloud or on premises or across other servers.

5.2 Semantic Metadata Services

Metadata can be defined as a set of assertions about things in our domain of discourse. We used Dublin Core metadata standard that can be used to represent properties or characteristics of resources[18].

We annotated the domain data using Dublin core standard metadata tags to create simple descriptive records for information to make data more meaningful and retrievable efficiently and inexpensively like Document title, date, time, creator.

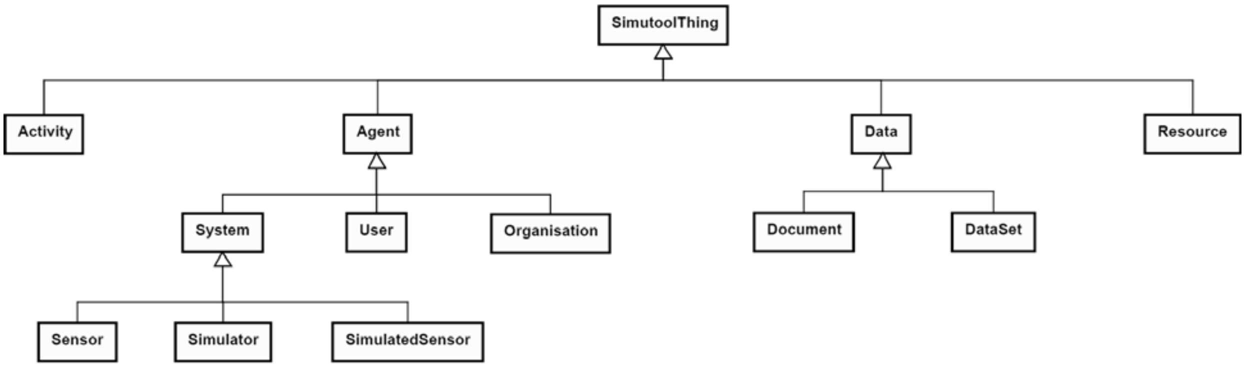

Fig. 2 Domain Model

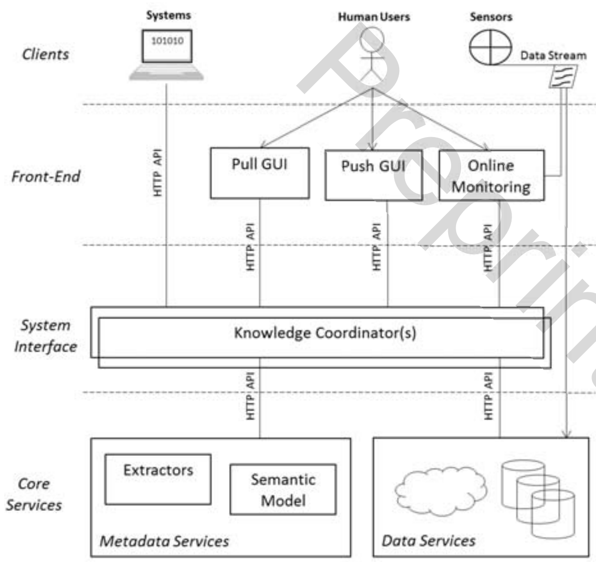

Fig. 3 Architecture of our KM system

5.2.1 Semantic Model

For semantic modeling we used Ontology Web Language (OWL) and RDF representation to develop the domain knowledge, RDF is a standard model for data interchange, RDFS is a language for representing simple RDF vocabularies and OWL as a computational logic-based language represents rich and complex knowledge about things, groups of things, and relations between things[11].

We used Jena API framework to extract data and write to RDF. Jena provides support for OWL and parses it into RDF [13]. The RDF representation of data is stored in a Triple Store. We used Apache Jena Fuseki server to perform the Triple store task and querying service. Fuseki is an HTTP server to RDF data. It is also a sub-project of Jena developed as a servlet and runs as a stand-alone server [20]. It supports SPARQL for querying and updating. SPARQL is an RDF query language and it is a semantic query language.

We used SPARQL query service to query and update the RDF Triple store. SPARQL is able to retrieve and manipulate data stores in RDF format[8].

5.3 Data Services

The manufacturing environment produces various digital artifacts which can be categorized under one of documents or data. Data is structured and machine readable, documents may not machine readable or *not intended* to be used that way; they are stored to be produced for human users to explore, edit, and store them. Furthermore, due to the multi-disciplinary nature of the experimental manufacturing domain, data in various formats and natures are produced each of which is best stored and managed using different data management systems. Take the curing process simulation data for instance, due to the large volume and fine granularity of the data, it is best stored in either a time-series database system such as influxDB[9] or OPENTSDB[10] or in an NOSQL system with a schema designed to fit the usage requirements, and that is only one type of data. To tackle these issues we have developed the data service component to tackle several of these issues.

Figure 4 shows an outline of the internals of the data services section, demarcated by a dashed line, with technology examples. It can be divided into three parts, each will be discussed below.

5.3.1 Data Custodian

In the traditional data governance, there are usually two overlapping duties: Data Stewards [30] and Data

[9] `https://www.influxdata.com`
[10] `http://opentsdb.net`

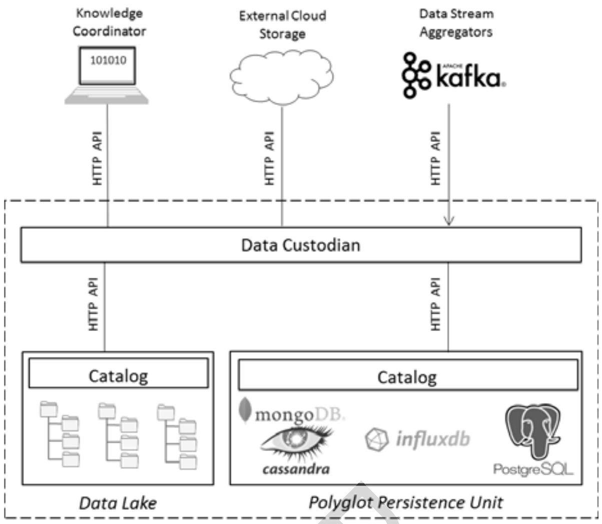

Fig. 4 Detailed View of the Data Services

Custodians [29,1]. Whereas stewards are responsible for data content, context and business rules, Data Custodians take a cue from the traditional data governance literature and adapt them to digital data management. Data custodian is responsible for the operational and technical environment of the data management. The main responsibilities of the data custodian in our work is outlined below.

- **Security**. The location of the data custodian in the architecture allows it to focus on certain level of security in data access and alleviates some of these responsibilities from the storage systems.
- **Locating Storage Sinks**. Given new artifacts to store, identify the correct storage sink to forward them to. On the other hand, given a request for accessing a digital artifact, identify where to forward that request to. We use MIME types[11] (Multipurpose Internet Mail Extensions) to communicate the type of the digital artifact. This is achieved by adding a metadata tag to in semantic model discussed in the previous section. Later on, rules can be written to handle different kinds of resources, here is an example rule. Take MP4 videos, which which we consider as documents (intended to be stored and served back to human users or external systems). The MIME type of MP4 videos is `video/vnd.uvvu.mp4`, so it will be attached as metadata to the digital artifact. Hence a simple rule here will be that `video/vnd.uvvu.mp4` must be forwarded to the data lake. MIME types can be automatically extracted by extractors, and rules can be written for them as well[12].

5.3.2 Polyglot Persistence Unit

Polyglot Persistence [24,25] is the use of more than one data storage service in a single application. As mentioned in the start of this section, data storage requirement of disparate data formats and domains are rarely met in one data storage solution. We use the Polyglot Persistence Unit for storing structured data that we our our partners systems need to process, query and carry out computations on. The Polyglot Persistence System is divided into two parts: the Catalog and the data management systems. The Catalog is responsible for providing inventory services, as well as data availability; it is at the right level to accommodate replication of data sources seamlessly.

To uniquely identify and reference data-sets or data-artifacts, we use an identification system to uniquely identify each data-set or data-artifact, to ease in referencing and accessing stored data. This identification is also added as a field into our semantic metadata model. Currently, we use a query to identify a data-set, but we are also investigating other methods to identify to identify data-artifacts in general.

5.3.3 Selection of Data Management Systems

Regarding the data management systems, we currently use an RDBMS (we use PostgreSQL[13]) and we are benchmarking and comparing NOSQL systems (document and column stores) and time-series databases to understand their strengths and weaknesses in the experimental manufacturing domain. Times series databases seems like a good fit for simulation data, but we are also interested to see the efficiency and usability of NOSQL systems when their models are designed to handle the uniques requirements of a domain such as manufacturing. This last part is one of our current research interests, particularly, whether the specificity of the domain can be used to build data management systems that clearly outperform generic solutions, and what is the utility in such decisions.

As an example of the specificity of the digital manufacturing domain, consider activities such as production on-line monitoring (see Chapter 6) and the identification of faulty and successful production runs; these are key requirements from the data management side and

[11] `https://developer.mozilla.org/en-US/docs/Web/HTTP/Basics_of_HTTP/MIME_types`

[12] We use Apace Tika (`https://tika.apache.org/`) as our extractor it is an open source, easily extendable framework of extracting meta-data and content of digital files.

[13] `https://www.postgresql.org/`

they have their peculiar challenges, but also simplify other aspects of the data management.

5.3.4 Data Lake

The Data Lake is where documents are stored, as well as source versions of data the system parsed and stored in its data warehouse. We use a file system as the storage model for the Lake. There can be many reasons to keep the source data formats, it can be for lineage reasons, or that some clients of the KBMS system needs the data in original. Take for example simulation data outputs. Although the contents of these files is parsed into the data warehouse, some systems or the original unloader might need the original file to load it using the simulation software[14]. A data lake is as good as its catalog and inventory services are. Although we use the similar concept of the Catalog in both the Data Lake and Polyglot Persistent Unit, we are interested in finding out factor out a core of the catalog to serve both storage domains. Currently we use a simple file-system for the data lake, and we are still investigating which features and functionality do we require out of Data Lake catalog.

6 Online Monitoring

With the availability of simulation data, from the past, and sensor data from the present during production and experiments, one of the requirements of the domain is to compare all this data online, during the production process. The goal of it is to support the line engineers and enrich then with insights and experiences from past simulations and runs. The KBMS already provides a framework to store and service all this data, along with a set of interface to interact and interface with it. What remains is an application that plugs into it and provides the functionality of online monitoring. See Figure 3 in Chapter 5 to get an idea where the online monitoring component fits within the KBMS ecosystem. In this chapter we discuss some ideas we have developed for the online monitoring system.

For the monitoring system there are lot of domains where sensor based approaches are used to adapt the same approach as we are using for the SIMUTOOL knowledge management system (KBMS), e.g. to use a wireless sensor network to solve the climate control problem in a greenhouse [21].

[14] A policy of what to keep in original format and what to throw away is partly domain dependent but it is future work to investigate whether general patterns can be identified

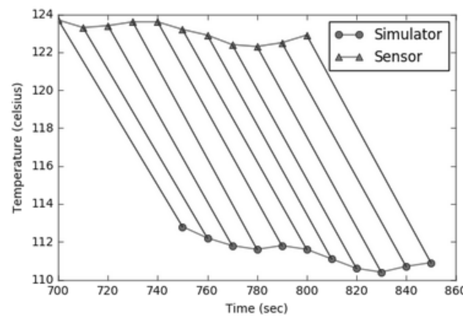

Fig. 5 Window Concept Past

Our monitoring system evolves in the experiment and production phase (see Figure 5). In the experiment phase the monitoring system is used to adjust the different settings and to learn how the production of the components is achieved in a successful way. In the production phase the monitoring system is more useful in the control aspect to detect variation between the results in the simulation and the production. The concrete realization will be realized in the online monitoring - a component to monitor live sensor data and compare it with simulation data (see Figure 5).

The sensor and the simulation data are available as time series data. Time series data are a set of data points in temporal order. To compare time series data, Dynamic Time Warping (DTW) is used. DTW is an algorithm for measuring similarity between two temporal sequences which may vary in speed. To extend the strengths of DTW, we use window concepts to check if the sensor data is delayed in relation to the simulation data. This is realized with a present window where the data points in the same time range are compared, and two outlook windows where the simulation data is shifted to the future.

The sensor data could be a little bit delayed or to fast in relation to the stored simulation data. To use the strength of the DTW algorithm we introduce an additional window concept as shown in Fig 5. The window concept simply consists of the usage of three different time windows. In the following the parameters T equals to the time now and W is the used window size. The three time windows are:

- Present: [T-W; T]
- Past: [T-2*W; T-W]
- Future: [T; T+W]

We discussed our proposals of the online monitoring system with the end-users form the production side and have received positive feedback, with more ideas and suggestions to adopt the online monitoring to the production environment. It also serves as an example application of the KBMS.

7 Discussion & Conclusion

we showed how experimental manufacturing processes can be supported by semantic data management. By modeling the processes and building a domain model of the domain for clear communication as well as for building the semantic layer atop of the data services, as well as a extending the annotation system based on changing activities beyond the Dublin Core metadata scheme. We have also shown how polyglot persistence can be achieved within the context of the KM and how it can fit smoothly supporting the operation of the KM. Showing how a modular and service-oriented architecture can be easily extendable by application programs and external systems, with the online monitoring as an example of such a system. For our future work, we would like to investigate in scalable data services, adaptive UI for different activities and roles, adding more extractor, and transferring the work we do into other domains, e.g., managing large-scale sensor-based infrastructure of a smart city, which is another active topic of research we are pursuing [27].

References

1. REG 08.00.03 Data Management Procedures Policies, Regulations & Rules. URL `https://policies.ncsu.edu/regulation/reg-08-00-03/`
2. Alavi, M., Leidner, D.E.: Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MIS Quarterly **25**(1), 107–136 (2001). DOI 10.2307/3250961. URL `http://www.jstor.org/stable/3250961`
3. Boudreau, M.C., Loch, K.D., Robey, D., Straud, D.: Going Global: Using information technology to advance the competitiveness of the virtual transnational organization. Academy of Management Perspectives **12**(4), 120–128 (1998). DOI 10.5465/AME.1998.1334008. URL `http://amp.aom.org/cgi/doi/10.5465/AME.1998.1334008`
4. Buneman, P., Chapman, A., Cheney, J.: Provenance Management in Curated Databases. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, SIGMOD '06, pp. 539–550. ACM, New York, NY, USA (2006). DOI 10.1145/1142473.1142534. URL `http://doi.acm.org/10.1145/1142473.1142534`
5. Cavalcante, E., Alves, M.P., Batista, T., Delicato, F.C., Pires, P.F.: An Analysis of Reference Architectures for the Internet of Things. pp. 13–16. ACM Press (2015). DOI 10.1145/2755567.2755569. URL `http://dl.acm.org/citation.cfm?doid=2755567.2755569`
6. Chervenak, A., Foster, I., Kesselman, C., Salisbury, C., Tuecke, S.: The Data Grid: Towards an architecture for the distributed management and analysis of large scientific datasets. J. Netw. Comput. Appl. **23**(3), 187–200 (2000). DOI 10.1006/jnca.2000.0110. URL `http://dx.doi.org/10.1006/jnca.2000.0110`
7. Civili, C., Console, M., De Giacomo, G., Lembo, D., Lenzerini, M., Lepore, L., Mancini, R., Poggi, A., Rosati, R., Ruzzi, M., others: MASTRO STUDIO: managing ontology-based data access applications. Proceedings of the VLDB Endowment **6**(12), 1314–1317 (2013). URL `http://dl.acm.org/citation.cfm?id=2536304`
8. Coppens, S., Sande, M.V., Verborgh, R., Mannens, E., de Walle, R.V.: Reasoning over SPARQL. In: Linked Data on the Web 2013., vol. 996. CEUR-WS, Rio de Janeiro, Brazil (2013). URL `http://ceur-ws.org/Vol-996/papers/ldow2013-paper-08.pdf`
9. Daraio, C., Lenzerini, M., Leporelli, C., Naggar, P., Bonaccorsi, A., Bartolucci, A.: The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality. Scientometrics **108**(1), 441–455 (2016). DOI 10.1007/s11192-016-1913-6. URL `https://link.springer.com/article/10.1007/s11192-016-1913-6`
10. Davenport, T.H., De Long, D.W., Beers, M.C.: Successful Knowledge Management Projects. Sloan Management Review; Cambridge, Mass. **39**(2) (1998). URL `https://search.proquest.com/docview/1302989262?pq-origsite=gscholar`
11. Dean, M., Schreiber, G.: OWL Web Ontology Language reference (2004). URL `https://www.w3.org/TR/2004/REC-owl-ref-20040210/`
12. Gray, J., Liu, D.T., Nieto-Santisteban, M., Szalay, A., DeWitt, D.J., Heber, G.: Scientific Data Management in the Coming Decade. SIGMOD Rec. **34**(4), 34–41 (2005). DOI 10.1145/1107499.1107503. URL `http://doi.acm.org/10.1145/1107499.1107503`
13. Jeek, P., Mouek, R.: Semantic framework for mapping object-oriented model to semantic web languages. Frontiers in Neuroinformatics **9**, 3 (2015). DOI 10.3389/fninf.2015.00003. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4340193/`
14. Kharlamov, E., Bilidas, D., Hovland, D., Jimenez-Ruiz, E., Lanti, D., Lie, H., Rezk, M., Skjaeveland, M., Soylu, A., Xiao, G., Zheleznyakov, D., Giese, M., Ioannidis, Y., Kotidis, Y., Koubarakis, M., Waaler, A.: Ontology Based Data Access in Statoil. Web Semantics: Science, Services and Agents on the World Wide Web **44** (2017). URL `http://www.websemanticsjournal.org/index.php/ps/article/view/497`
15. Kharlamov, E., Kotidis, Y., Mailis, T., Neuenstadt, C., Nikolaou, C., zep, ., Svingos, C., Zheleznyakov, D., Brandt, S., Horrocks, I., Ioannidis, Y., Lamparter, S., Mller, R.: Towards Analytics Aware Ontology Based Access to Static and Streaming Data. In: The Semantic Web ISWC 2016, Lecture Notes in Computer Science, pp. 344–362. Springer, Cham (2016). DOI 10.1007/978-3-319-46547-0_31. URL `https://link.springer.com/chapter/10.1007/978-3-319-46547-0_31`
16. Lee, S.M., Hong, S.: An enterprisewide knowledge management system infrastructure. Industrial Management \& Data Systems **102**(1), 17–25 (2002). DOI 10.1108/02635570210414622. URL `https://doi.org/10.1108/02635570210414622`
17. Li, Y.F., Kennedy, G., Davies, F., Hunter, J.: PODD - Towards an Extensible, Domain-Agnostic Scientific Data Management System. pp. 137–144. IEEE, Brisbane, QLD, Australia (2010). DOI 10.1109/eScience.2010.44. URL `http://ieeexplore.ieee.org/document/5693910/`

18. M.Y, C.: Perceived Helpfulness of Dublin Core Semantics: An Empirical Study. pp. 135–145. Springer, Cham (2013). DOI 10.1007/978-3-319-03437-9_14. URL https://link.springer.com/chapter/10.1007/978-3-319-03437-9_14
19. O'Leary, D.: Enterprise knowledge management. Computer **31**(3), 54–61 (1998). DOI 10.1109/2.660190. URL http://ieeexplore.ieee.org/document/660190/
20. Panchal, R., R.Swaminarayan, P.: Execution of SPARQL Query using Apache Jena Fuseki Server in AISHE domain. International Journal of Advance Engineering and Research Development **4**(9), 387–394 (2017). URL ExecutionofSPARQLQueryusingApacheJenaFusekiServerinAISHEdomain
21. Pawlowski, A., Guzman, J.L., Rodrguez, F., Berenguel, M., Snchez, J., Dormido, S.: Simulation of Greenhouse Climate Monitoring and Control with Wireless Sensor Network and Event-Based Control. Sensors **9**(1), 232–252 (2009). DOI 10.3390/s90100232. URL http://www.mdpi.com/1424-8220/9/1/232/
22. Poggi, A., Lembo, D., Calvanese, D., Giacomo, G.D., Lenzerini, M., Rosati, R.: Linking Data to Ontologies. In: Journal on Data Semantics X, Lecture Notes in Computer Science, pp. 133–173. Springer, Berlin, Heidelberg (2008). URL https://link.springer.com/chapter/10.1007/978-3-540-77688-8_5. DOI: 10.1007/978-3-540-77688-8_5
23. Rodrguez-Gonzlez, A., Stantchev, V., Colomo-Palacios, R.: Special issue on exploiting semantic technologies with particularization on linked data over grid and cloud architectures. Future Generation Computer Systems **32**(C), 260–262 (2014). DOI 10.1016/j.future.2013.10.021. URL https://dl.acm.org/citation.cfm?id=2748374
24. Sadalage, P.J., Fowler, M.: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, 1st edn. Addison-Wesley Professional (2012)
25. Schaarschmidt, M., Gessert, F., Ritter, N.: Towards Automated Polyglot Persistence. Workshop Paper, BTW (2015)
26. Sengstock, C., Mathis, C.: SAP HANA Vora: A Distributed Computing Platform for Enterprise Data Lakes. In: BTW, p. 521 (2017)
27. Steuer, S., Benabbas, A., Kasrin, N., Nicklas, D.: Challenges and Design Goals for an Architecture of a Privacy-preserving Smart City Lab. Datenbank-Spektrum **16**(2), 147–156 (2016). DOI 10.1007/s13222-016-0223-8. URL https://link.springer.com/article/10.1007/s13222-016-0223-8
28. Tnne, A.: Big Data is no longer equivalent to Hadoop in the industry. In: BTW, pp. 523–524 (2017)
29. University, C.M.: Data Custodians - Information Security Office - Computing Services - Carnegie Mellon University. URL http://www.cmu.edu/iso/governance/roles/data-custodian.html
30. University, C.M.: Data Stewards - Information Security Office - Computing Services - Carnegie Mellon University. URL http://www.cmu.edu/iso/governance/roles/data-steward.html