Institute of Architecture of Application Systems

Industrial Data Sharing with Data Access Policy

Felix W. Baumann[1], Uwe Breitenbücher[2], Michael Falkenthal[2],
Gerd Grünert[1], Sebastian Hudert[1]

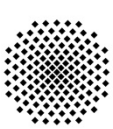
[1]TWT GmbH Science & Innovation
Stuttgart, Germany
[firstname].[lastname]@twt-gmbh.de

[2]Institute of Architecture of Application Systems,
University of Stuttgart, Germany
[lastname]@iaas.uni-stuttgart.de

BibTeX:

```
@conference{Baumann2017,
   author = {Baumann, Felix W. and Breitenb{\”u}cher, Uwe and Falkenthal, Michael
   and Gr{\”u}nert, Gerd and Hudert, Sebastian},
   doi = {10.1007/978-3-319-66805-5},
   booktitle = {Proceedings of the 14\textsuperscript{th} International Conference
   on Cooperative Design, Visualization, and Engineering (CDVE 2017)},
   title = {{Industrial Data Sharing with Data Access Policy}},
   year = {2017},
   pages = {215--219}
}
```


The final publication is available at Springer via
http://dx.doi.org/10.1007/978-3-319-66805-5



University of Stuttgart
Germany

Industrial Data Sharing with Data Access Policy

Felix W. Baumann[1], Uwe Breitenbücher[2], Michael Falkenthal[2], Gerd Grünert[1], and Sebastian Hudert[1]

[1] TWT GmbH Science & Innovation, Industriestr. 6,
70565 Stuttgart, Germany,
[firstname].[lastname]@twt-gmbh.de
[2] Institute of Architecture of Application Systems,
University of Stuttgart,
70569 Stuttgart, Germany,
[lastname]@iaas.uni-stuttgart.de

Abstract. In current industrial settings, data is dispersed on numerous devices, systems and locations without integration and sharing capabilities. With this work, we present a framework for the integration of various data sources within an industrial setting, based on a mediating data hub. Within the data hub, data sources and sinks for this industrial application are equipped with data usage policies to restrict and enable usage and consumption of data for shared analytics. We identify such policies, their requirements and rationale. This work addresses an industrial setting, with manufacturing data being the primary use-case. Requirements for these policies are identified from existing use-cases and expert domain knowledge. The requirements are identified as reasonable via examples and exemplary implementation.

Keywords: Industrial Data, Data Aggregation, Policies, Data Hub

1 Introduction

Industrial data utilisation and usage is currently influenced by a number of domains such as Cloud Computing, the Internet of Things (IoT), smart services and smart data analytics [2], artificial intelligence, machine learning, and data mining. The previous concepts and technologies are all part of the fourth industrial revolution, called Industry 4.0 [6]. One commonality of these concepts is the increased reliance and foundation in data. Industrial settings and especially manufacturing enterprises create and consume large amounts of data from numerous data sources and sinks.

While endeavours in the field of Industry 4.0 are promising approaches to provide new insights into and to create opportunities from the analysed data and the underlying processes, many problems arise. Formerly isolated data sources are integrated which can cause compliance issues, privacy, security or even legal violations. Moreover, often business critical data and details about processing steps and whole production processes are to be analysed by data scientists, which are often externals to the data-owning companies and, furthermore, are currently

rare. This is due to the fact that the knowledge and expertise about analytics algorithms, techniques and platforms is typically not part of the core business of manufacturing companies. Further problems arise from the heterogeneity of data sources that must be unified and adapted for efficient and shared usage.

On the one hand, this trend is mainly driven by developments in IoT, allowing devices of reduced size and price. This miniaturization facilitates bringing out many sensors in manufacturing environments to collect data about production processes, processing steps of machinery, and surrounding parameters. Environmental parameters include humidity, light irradiation, and temperature in production environments [3, 10] and many more. On the other hand, cloud technologies and evolution of new analytical approaches and platforms enable the rapid processing of the acquired large datasets, even on-line via streaming analytics frameworks. Analytics platforms, such as Apache Flink [11], are developed under the constraint to be highly optimized to provide application programming interfaces and libraries specifically for developing analytics algorithms among runtimes and integration middleware. Flink allows data to be either processed via batch jobs or continuous data streams. Such analytics platforms and algorithms often profit from processing data in parallel, distributed among dynamically allocated compute nodes. Thus, they can leverage the scaling capabilities of cloud infrastructures, platforms and services, be it in public, hybrid or private clouds.

With this work, we propose a data integration and sharing framework that is constrained by a set of policies to enable the secure and efficient usage of distributed data sources within industrial environments. New optimization opportunities in manufacturing processes are leveraged by integrating data from a manifold of different data sources to overcome their isolation and enable holistic analysis approaches.

2 Related Works

Yu et al. [12] provided a rationale for the sharing of information amongst business partners, especially within a supply chain to minimise risks and uncertainties. In the context of clinical data, Malin et al. [8] discussed requirements, such as privacy, for data sharing to achieve beneficial results. These authors identified regulatory and legal constructs as restrictive functions. Gardner et al. [7] also researched the academic and medical domain of data sharing with mandatory requirements for specific cases and distinguished forms and methods of sharing, such as direct, i.e. two party, and public sharing. The work by Zhao et al. [13] on the secure data sharing over untrusted cloud storage providers also influenced our work, since issues of transitivity of rights are discussed therein. Breitenbücher et al. [5, 4] showed how policies can influence the deployment of applications, e.g., to enforce secure passwords or deployment in specific regions. Current work on the issue of data sharing is mainly focused on scientific data sharing, thus, only partially applicable for our industrial setting and, furthermore, does not explicitly combine the multitude of problems encountered, such as privacy and compliance awareness, security and heterogeneity of data.

3 Secure Data Integration and Sharing Framework

To enable the collaborative usage of data among partially competing entities, trust is required in safekeeping of information and enforcement of rules or policies. With this work, we propose such a trusted instance in form of policy enforcement directly at the logical location of the data source or sink. We present a secure data integration and sharing framework, that enables that every such source or sink is equipped with a filtering and access software component. This component is under the direct control of the respective data owner.

Finally, sharing data for enabling analytics approaches among many data sources can ultimately be extended to scenarios where formerly classified data is shared with external companies in an aggregated and obfuscated form. Thus, business secrets remain protected while new analytics opportunities are generated. Each of these issues demand that data security and privacy have to be assured to protect the data from illegitimate and undefined uses.

See Fig. 1 for a depiction of the implementation schematics. In this figure, the data hub is shown as the central rectangle that allows access to the four depicted data-sources and sinks (S1–S4), which can be of diverse type such as databases, machine and sensor data or file data, for authorised parties. The access is mediated through the triangular software adapters on premise of the data owner. The policies (indicated as P1–P4) are directly attached to the adapters and under the control of the data owner. These policies, of which there can be multiple for each source, are propagated from the source to the consumer or user, as data hubs can function as data sources and sinks for further data hubs, thus, allowing for propagated access. In the figure, Party A and Party B, both make use of the data hub. Both parties can be distinct and from different entities, with different properties of ownership of the data sources. The proposed framework is comprised of the data hub, the corresponding adapters and the policy enforcement component. The geometrical shapes in the figure indicate the variety of different data source types.

We implement adapters for various data sources and sinks that transform data to be uniformly accessed through the data hub. The data from the data hub is exposed through the OData protocol (OASIS Open Data Protocol [9]). This protocol enables third party software to interact with the data hub and its exposed data in an uniform and standardised method. The data hub integrates the varying schemas of the data sources so that unified querying and application of policies is enabled. Constraint and policy application is enforced at each point equipped with a policy.

We present in this work findings from the project SePiA.Pro [1], which investigates these issues in the above described context. Furthermore, we illustrate the elaborated requirements for protecting industrial data in the context of Industry 4.0 endeavours via data policies. Such data policies are means to specify constraints, restrictions, or instructions that apply to the data, taking into account aspects such as data accessibility, utilisation, processing, obfuscation, storage or generation. The policies extend common access control rules and restrictions to incorporate concepts such as temporal, logical, and organisational triggers.

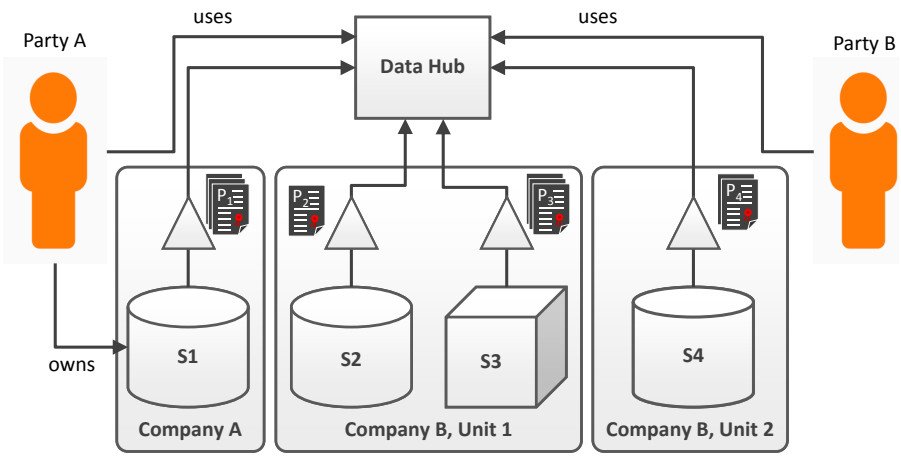


Fig. 1. Framework Architecture Overview

The policy definition is flexible and extensible to allow individual and specific policies to be defined by implementing parties. It is further discussed, how and through which means, i.e. systems and parties, such policies are enforced at several points in time of the lifecycle of smart services — specifically at modelling time, deployment time, and runtime — to overcome the above mentioned obstacles. Specific scenarios for enabling trust and enforcing implementation are analysed within this work. We also discuss the concept of attaching data policies to relevant data sources. The rationale for such an attachment of policies is to secure and protect data from manufacturing environments in standards-based deployment models such as cloud computing. These models are used to provision smart services and wiring them with arbitrary data sources, such as databases, data aggregation services, industry specific machine to machine or IoT related data streaming endpoints. We provide and discuss exemplary policies, such as the restriction of data consumption within specific premises or logical groupings within enterprises.

4 Summary

In this work, the rationale for data sharing components is provided. By attaching policies to data sinks and sources we have provided a method to enforce requirements for data processing for all involved parties. The parties are shown to keep sovereignty over their respective data, thus, potentially increasing the acceptance of collaborative data usage. It was shown, that such data usage can enable the creation of future smart services without the risk of unintentionally exposing sensitive data to unauthorised parties. We have shown, that the data hub as a central component for such shared data usage, can enable secure, privacy and compliance aware collaboration on data.

Acknowledgments

This work is partially funded by the project SePiA.Pro (01MD16013F) of the BMWi program Smart Service World.

References

1. Service Plattform für die intelligente Anlagenoptimierung in der Produktion, http://projekt-sepiapro.de, last accessed on 12th May 2017
2. Allmendinger, G., Lombreglia, R.: Four strategies for the age of smart services. Harvard Business Review 83(10), 131 (2005)
3. Atzori, L., Iera, A., Morabito, G.: The internet of things: A survey. Computer Networks 54(15), 2787–2805 (2010)
4. Breitenbücher, U., Binz, T., Fehling, C., Kopp, O., Leymann, F., Wieland, M.: Policy-Aware Provisioning and Management of Cloud Applications. International Journal On Advances in Security 7(1&2) (2014)
5. Breitenbücher, U., Binz, T., Kopp, O., Leymann, F., Wieland, M.: Policy-Aware Provisioning of Cloud Applications. In: Proceedings of the Seventh International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2013). pp. 86–95. Xpert Publishing Services (2013)
6. Dais, S.: Industrie 4.0 – Anstoß, Vision, Vorgehen, pp. 261–277. Springer Berlin Heidelberg, Berlin, Heidelberg (2017)
7. Gardner, D., Toga, A.W., et al.: Towards effective and rewarding data sharing. Neuroinformatics 1(3), 289–295 (2003)
8. Malin, B., Karp, D., Scheuermann, R.H.: Technical and Policy Approaches to Balancing Patient Privacy and Data Sharing in Clinical and Translational Research. Journal of Investigative Medicine 58(1), 11–18 (2015), http://jim.bmj.com/content/58/1/11
9. OASIS: Oasis open data protocol (odata). Tech. rep., OASIS (2014), http://docs.oasis-open.org/odata/odata/v4.0/os/part1-protocol/odata-v4.0-os-part1-protocol.html
10. Sundmaeker, H., Guillemin, P., Friess, P., Woelffle, S.: Vision and challenges for realising the internet of things. European Commission Information Society and Media (2010)
11. The Apache Software Foundation: Apache Flink: Scalable Stream and Batch Data Processing, https://flink.apache.org, last accessed on 12th May 2017
12. Yu, Z., Yan, H., Cheng, T.E.: Benefits of information sharing with supply chain partnerships. Industrial Management & Data Systems 101(3), 114–121 (2001)
13. Zhao, G., Rong, C., Li, J., Zhang, F., Tang, Y.: Trusted Data Sharing over Untrusted Cloud Storage Providers. In: 2010 IEEE Second International Conference on Cloud Computing Technology and Science. pp. 97–103 (11 2010)