FactDAG: Formalizing Data Interoperability in an Internet of Production

Lars Gleim, Jan Pennekamp, *Student Member, IEEE*, Martin Liebenberg, Melanie Buchsbaum, Philipp Niemietz, Simon Knape, Alexander Epple, Simon Storms, Daniel Trauth, Thomas Bergs, Christian Brecher, Stefan Decker, Gerhard Lakemeyer, Klaus Wehrle, *Member, IEEE*

***Abstract*—In the production industry, the volume, variety and velocity of data as well as the number of deployed protocols increase exponentially due to the influences of IoT advances. While hundreds of isolated solutions exist to utilize this data, e.g., optimizing processes or monitoring machine conditions, the lack of a unified data handling and exchange mechanism hinders the implementation of approaches to improve the quality of decisions and processes in such an interconnected environment.**

The vision of an *Internet of Production* promises the establishment of a *Worldwide Lab*, where data from every process in the network can be utilized, even interorganizational and across domains. While numerous existing approaches consider interoperability from an interface and communication system perspective, fundamental questions of data and information interoperability remain insufficiently addressed.

In this paper, we identify *ten* key issues, derived from three distinctive real-world use cases, that hinder large-scale data interoperability for industrial processes. Based on these issues we derive a set of *five* key requirements for future (IoT) data layers, building upon the FAIR data principles. We propose to address them by creating *FactDAG*, a conceptual data layer model for maintaining a provenance-based, directed acyclic graph of facts, inspired by successful distributed version-control and collaboration systems. Eventually, such a standardization should greatly shape the future of interoperability in an interconnected production industry.

***Index Terms*—Data Management, Data Versioning, Interoperability, Industrial Internet of Things, Worldwide Lab.**

I. Introduction

Today, the usage of large scale *Cyber-Physical Production Systems* (CPPS) is more and more applicable to productive environments as concepts are maturing [1]. In a nutshell, CPPS digitize analog processes, e.g., manufacturing processes, by applying sensor based data acquisition approaches followed by data storing and processing concepts known from computer science [2]. Research in CPPS also focuses on methods to integrate highly heterogeneous data sets and mixtures of different communication protocols that are used by the involved machines [3]. The ever-growing amount of data in the industry – due to the usage of sensor-systems and interconnected machines in an *Internet of Things* (IoT) – needs to be transmitted, stored, and processed for each machine connected in the system [4]. This progression results in an increase of the total amount of data at an exponential rate [5], [6].

Therefore, one goal should be to analyze the collected product data and utilize it in (sophisticated) models to improve the currently applied production processes. These advancements can be accelerated, if the usage of data from different companies, manufacturing plants, and machines equipped with different tools is incorporated as well [7]–[9]. However, any improvements are only realized, if process data is widely available and the information from various sources can be integrated with reasonable effort. Currently data is typically either (i) not stored at all, (ii) only retained locally, i.e., stored in isolated data silos, or (iii) not transferable between systems by different manufacturers due to vendor-specific solutions (as identified by [10]).

In an interdisciplinary research cluster at RWTH Aachen University, we plan to tackle the aforementioned problems with the idea of an *Internet of Production* (IoP) [11]. A central goal of an IoP is to establish a *Worldwide Lab* (WWL) which enables automated exchanges of process data across domains, organizational structures, and companies, promising to unlock unrealized synergies through digitized collaboration [12]. The idea is that we consider every production step in the WWL as an experiment so that processes on the other side of the world can be improved with the knowledge gained from these experiments. A fundamental requirement for the realization of such experiments is the capability to continuously pool data from various manufacturing processes, potentially originating from different companies with widely varying implementations and protocols. Therefore, in a WWL and thus in an IoP, the need for an holistic data interoperability layer exists.

Such a layer connects the major domains of a producing company (i.e., product development, production technology, production planning). Thus, enabling cross-domain collaboration by providing (semantically) adequate and context-aware data from production, development and usage, even in real-time and across organizational boundaries, ultimately shaping an interconnected production industry. An interoperability layer across vendor specific solutions is a key building block

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612.

L. Gleim* and S. Decker are with the Chair of Databases and Information Systems. J. Pennekamp and K. Wehrle are with the Chair of Communication and Distributed Systems. M. Liebenberg and G. Lakemeyer are with the Knowledge-Based Systems Group. M. Buchsbaum, P. Niemietz, S. Knape, A. Epple, S. Storms, D. Trauth, T. Bergs, and C. Brecher are with the Laboratory of Machine Tools and Production Engineering. All authors are affiliated to RWTH Aachen University, Germany; S. Decker additionally to Fraunhofer FIT, Germany and T. Bergs to Fraunhofer IPT, Germany.

*Corresponding author email: gleim@cs.rwth-aachen.de

Manuscript received July 22, 2019; revised December 17, 2019.

towards overcoming one of the most important barriers to the establishment of advanced IoT approaches [13]. In this paper, we formalize the data layer model FactDAG as a groundwork for data interoperability in an IoP.

The remainder of this paper is structured as follows. In Section II, we analyze the state of data interoperability based on related work and three distinctive real-world use cases. Afterward, in Section III, we identify *ten* related key issues before deriving a set of *five* requirements for data interoperability in an IoP. We propose to address them by creating *FactDAG*, a data layer model, for maintaining a directed acyclic graph of immutable *Fact*s, persistently identified using *FactID*s. Further, in Section IV, we highlight the respective benefits and then, we conclude in Section V.

II. The State of Data Interoperability

To give an overview about current advances towards data interoperability, we present recent approaches and data models. Afterward, we analyze three typical use cases to understand their perspective wrt. their data needs. As a foundation, we summarize these findings to define a mutual perspective of data interoperability and derive respective existing key issues in a structured form.

A. Existing Approaches Towards Data Interoperability

Well-established technologies, such as *Cloud Computing* [14] and *IoT* [15], were introduced into production systems during the last years. In this process, collaboration between production engineers, computer scientists and technicians from related fields have grown increasingly closer, driving the need for cross-domain data and system interoperability. Existing approaches strive to enable interoperability between IoT protocols and platforms as well as in development toolchains used to design and execute manufacturing processes. In recent years, multiple architectures, frameworks, and layers for interoperability, including semantic approaches, were introduced [16]–[20]. Notable examples include OMG Data Distribution Service (DDS) [21], MTConnect [22], OPC Unified Architecture (OPC-UA) [23], the BIG IoT Platform [16] and the Industrial Data Space (IDS) initiative [24]. Typically covered functionalities include authentication, registration and discovery, as well as accounting of and access to resources across IoT platforms.

Another perspective describes the usage of behavior- and domain-driven development in a framework, which implicitly outputs a domain model that serves as a foundation for semantic interoperability [19]. While Nilsson et al. [20] give an overview of semantic driven approaches, Gürdür et al. [25] examine interoperability regarding toolchains for measurement systems. They conclude that introduced interoperability assessment models focus on different aspects regarding interoperability and due to a lack of comprehensiveness, these models are not used in industrial contexts.

While all approaches provide important input for (data) interoperability in the IoT, the complex multi-party and multi-step collaboration scenarios of the manufacturing environment can still not be modeled with sufficient clarity for collaboration and data reuse. Especially when data is shared across organizational- or domain-boundaries, implicit context knowledge and data provenance is regularly lost, reducing the interpretability of data.

One explanation for the lack of data interoperability may be that a dominant approach towards information processing and exchange in CPPS is the classical extract, transform, load (ETL) procedure, which was first described in 1977 [26] and later popularized with the emergence of data warehousing systems [27], that only focuses on unidirectional dataflows from source to sink and typically does not regard the challenge of interorganizational information exchange in any specific way.

B. Interoperability Needs Arising from Case Studies

To motivate requirements for data interoperability, which allow for an improved information exchange, we further introduce three practical use cases in different production scenarios – practically coinciding e.g., in the supply chain of car manufacturing – which pose several challenges on data interoperability.

1) Mass-Production Process Monitoring: Fine blanking is an economical precision forming process to produce large number of identical workpieces commonly employed in the automotive industry [28]. Once set up correctly, the process of fine blanking is a stable process that constantly produces high quality workpieces [29]. Although the setup does not change throughout the process, none of the produced workpieces are identical [30] due to environmental conditions, tool wear, changing surface quality or inhomogenous material properties [31]. As a result, to fully understand the extent of important quality features and minimize the amount of rejected workpieces, simply monitoring the fine blanking process itself is insufficient, instead information of the used material acquired by its manufacturer have to be taken into account. With the availability of data given by the material manufacturer, the objective of a flexible process setup based on the specific material properties can be addressed. Taking the adaptivity of the process into account, an automatic adoption of the process accounting can react on changes in material properties or lubrication behaviour with the help of data-driven models [32]. However, as models giving recommendations for actions may cause damage to the machine tools or other equipment, the integrity and accountability of the underlying data gains more importance, especially due to the associated questions of liability for erroneous data. The development of models as well as their constant improvements in the sense of an WWL rely on the availability and traceability of data throughout the variable process chain, which requires systems to be interoperable at all times.

2) Process Optimization and Prediction: Due to the high variety of machine tools, materials and final workpiece requirements, choosing the right process parameters for fast and reliable processes is difficult. Today's machine tools have a huge variety of integrated sensors, generating a continuous stream of sensor data and process control data over time. By combining this data with knowledge from a company's central information repositories (e.g., ERP, MES or MRP systems) deep real-time insights of the current process can be achieved. For example, Königs et al. [33] developed a virtual workpiece model that matches the real machining process. However, to create, operate and validate these models, large amounts of data from different sources, even interorganizational, are required. To establish collaborations with the goal of utilizing process data, new forms of interoperability have to be developed wrt. controlled accessibility and data sovereignty. Furthermore, data exchanges between the different systems and stakeholders must be supported. Currently, most of the available information is stored in proprietary systems, which require individual interfaces to access data. The design of these proprietary systems exacerbates the situation as contextual information or even information about the source of the data can get lost.

3) Industrial Control of a Production Line: For example, as part of his production line, a glass manufacturer has quality issues because his product has high tolerances and due to thermal processes, the output varies during the day. The robots, which have to handle the glass, are not sufficiently able to react to changes in the surface geometry of the product. To tackle this challenge, a measurement system must be used to identify surface deviations and provide this information to each robot. All components of the production line are connected in an IoP on different layers (Cloud, Edge, Shopfloor) to support the information transfer which is ultimately collected and processed to reconfigure the currently running process. Production process information – such as the state of each component (e.g., proximity sensor, pneumatic, robot, camera), robot trajectories, etc. – has to be available, rapidly retrievable by all devices and services at all times, and in a consistent form, to establish a resilient process control [34].

For detecting process anomalies, as well as for establishing new business strategies involving the car manufacturer or the customer, production data must be made available through a platform providing information about the product and the production facilities, allowing to trace back data origins and influences. This data sharing enables data usage for process optimization and therefore, creates a feedback loop for the running system. Besides, due to the possibility of providing new services to customers, additional value is created for the car manufacturer. To establish reliable data sources and to facilitate accountability in such distributed collaboration scenarios, principles must be implemented that provide information of the data origin including an identifier, its authority, and the origin of changes, enabling the interpretation of the context, history, and reliability of used data. The high variety of tasks for running a production line - such as scheduling, controlling the process, providing product information and quality inspection - rely on the same data, i.e., use and write that data. Therefore, this data has to be made available in an interoperable manner.

C. Key Issues Motivating a Data Layer Model

The described use cases illustrate that interoperability between different departments and production sites as well as across company boundaries can increase productivity and unlock previously unrealized value. The adoption of tighter integrated production processes and data driven workflows is however currently hindered by a number of fundamental limitations. In the following, we identify *ten* key issues that are present in today's data layer models:

I1 *Data Inaccessibility*: Even though the presented use cases express different general needs (e.g., context and provenance information, data sovereignty, or accountability), all of them are constrained by the current situation where data is not accessible, only retained locally, or limited to a vendor-specific solution (cf. Section I). Accessing information across system boundaries and from different stakeholders is limited by incompatible interfaces in industry and proprietary systems, largely embracing vendor lock-in over interoperability. As such, data frequently resides in isolated data silos, inaccessible from the outside and may be considered as unrealized potential or hidden value [12].

I2 *Lack of a Clear Data Authority*: Currently, as illustrated in Section II-A, most implemented interoperability layers have a (limited) local view on the problem. As such, the authority under which data was created (i.e., the liable and responsible party) is often not explicitly recorded and subsequently unclear when sharing data [35].

I3 *Data Reliability and Dependability*: Closely related to origin of data is the question of data reliability and dependability, both in terms of data, as well as service quality, as unreliable sources can lead to severe disruptions and damages in conjunction with tightly coupled processes [36], [37].

I4 *Data Mutability* and **I5** *Lack of Persistent Identifiers*: Additionally, data (e.g., sensor readings and database state) are often mutable [38], [39] and thus can be overwritten by newer versions of that data and subsequently cannot be reliably and persistently referenced [40].

I6 *Lack of Proper Data Versioning*: Exasperating this point is, that it is often unclear when new versions of a given data point are available, which is the latest version or even which of two versions is the more recent revision [41].

I7 *Lack of Provenance Information*: Directly related is the frequent lack of information regarding data formation history, origins and influences – its so-called provenance information – often rendering it impossible to judge the quality, relevance and applicability of information [42].

I8 *Lack of Semantic Embedding*: Another challenge is the general loss of context information and embedding, i.e., the relation between different data, their semantics and related entities, that limit the interpretability and reusability of data and can lead to significant damages in case of misinterpretation [42].

I9 *Legal Insecurity*: From the data consumer's perspective, a further key concern is that legislation and license restrictions which apply are non-transparent, extending to questions of ownership and liability, including for any derivative works and the potential damages they incur [43].

I10 *Lack of Control Over Data*: Similarly, from the data provider's perspective, once data has been shared with another party, control over it is typically lost [43].

Consequentially, existing concepts lack an adequate support of data reuse (and sharing) due to the missing comprehensiveness, effectively limiting the advances in production technology even though a sufficient number of IoT enabling approaches exists. For deployments in modern visions such as an IoP, which proposes the creation of a WWL, a unified data layer model shared across organizational entities could help in alleviating current shortcomings.

III. Interorganizational Data Interoperability

The presentation of the status quo shows that current solutions are not yet able to meet all the identified desired aspects of data interoperability and derived corresponding key issues. Based on this list, we further derive a set of requirements that we address by presenting a data layer model for future data interoperability.

A. The Road Ahead

We are convinced, that a fundamental requirement for the realization of an interconnected network of production systems to create a WWL is the wide-spread (and preferably open) availability of an abstract interaction, interoperability and data exchange model, as well as a concrete interface implementation, which ensures that data is uniform, aggregable and semantically enriched while data access is properly permissioned and audited.

Permitting access to, aggregating and combining production data from different sources allows for individual data to be reused in different contexts. Additionally, even new business models can be created. For example, an analysis and generation of prediction models (cf. Section II-B) could be offered solely based on the accessible data. Similarly, an agile supply chain management structure could be established. A significant difference to today's potential is the combination of different data sources: Companies no longer only produce workpieces, but also data, which leads to the concept of WWLs. Hence, individual process data can be brought into different contexts to generate added value.

A goal within an IoP is the processing of data across locations and organizational boundaries [12]. Consequentially, potentially sensitive data is being transferred away from the owner's jurisdiction over to potentially untrusted third parties. Hence, an appropriate protection of sensitive information as well as intellectual property within such a system is necessary to meet any company's need for data confidentiality. Otherwise, any collaboration will likely fail due to conservative attitudes regarding data sharing [12], [43].

B. Requirements for (Internet-scale) Interoperability

To address the identified issues, we extend upon the FAIR data principles [44], as summarized in table I, and derive *five* corresponding additional requirements which would enable even sophisticated interoperability in the future (if properly implemented and deployed).

Initially devised in the context of scientific data management and stewardship to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows, the FAIR data principles are a set of guiding principles to enable data interoperability. As such, they already cover a number of the identified issues. Specifically, the directly addressed issues are **I1** (principles A1, A1.1, A1.2), **I5** (principle F1), **I7** (principle R1.2), **I8** (principles I2, I3, R1, R1.3), and **I9** (principle R1.1), as detailed in [44].

TABLE I
A summary of the FAIR data principles as specified in [44].

Findability	**Accessability**
F1 (meta)data are assigned a globally unique and eternally persistent identifier. **F2** data are described with rich metadata. **F3** (meta)data are registered or indexed in a searchable resource. **F4** metadata specify the data identifier.	**A1** (meta)data are retrievable by their identifier using a standardized communications protocol. **A1.1** the protocol is open, free, and universally implementable. **A1.2** the protocol allows for an authentication and authorization procedure, where necessary. **A2** metadata are accessible, even when the data are no longer available.
Interoperability	**Reusability**
I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. **I2** (meta)data use vocabularies that follow FAIR principles. **I3** (meta)data include qualified references to other (meta)data.	**R1** meta(data) have a plurality of accurate and relevant attributes. **R1.1** (meta)data are released with a clear and accessible data usage license. **R1.2** (meta)data are associated with their provenance. **R1.3** (meta)data meet domain-relevant community standards.

To accommodate for the remaining challenges, we derive the following corresponding data interoperability requirements:

RQ1 Data must be (a) immutable, (b) referenceable, and (c) uniquely identifiable (*Immutability, Referenceability, Unique Identifiability*) to address issues **I4** and **I5**.

RQ2 *Authoritative Source*: Data must be clearly attributed to a responsible authority. That authority is the singular authoritative source of the data (the single source of truth) and as such liable and responsible for the data. This aspect addresses **I2**, as well as potentially **I3**, whenever data reliability can be associated with the responsible authority.

RQ3 *Versioning:* Data must be explicitly versioned, providing a natural ordering of data revisions to address **I6**.

RQ4 *Data Sovereignty*: Data must be subject to definite legislation and usage regulations to address **I9** and **I10**.

RQ5 *Stakeholder Confidentiality* and *Minimalism*: Data access should only be granted to the absolute minimum of required information. Data sharing should occur in a privacy-respecting manner while applying authenticated encryption whenever reasonable. These requirements result from the industrial context to limit the possibility for loss of control over information to address **I10** and align well with the principle A1.2 of the FAIR data principles.

In the remainder of the paper, we will focus on the requirements **RQ1** to **RQ4** as they directly affect the way data is accessed, exchanged, and handled. In contrast, requirement **RQ5** mainly concerns the actual implementation of an interorganizational data layer model. Therefore, approaches must be in place to make sure that information is properly secured (stakeholder confidentiality). At the same time, the decision what scope of information corresponds to the minimally required amount of data (minimalism) can only be made domain-dependent, i.e., it is not part of the model itself either. Consequentially, this requirement can only be addressed after the underlying data interoperability model has been established. Thus, in the following, we derive an actionable blueprint for data interoperability in industrial scenarios.

C. From Data to Facts

In the following, we refer to data fulfilling requirements **RQ1** and **RQ2** – i.e., data that is immutable, referenceable, uniquely identifiable, and attributed to an authoritative source – as **facts**. Since facts (e.g., a distinctive physical sensor reading) do not change, they can reliably be referenced and built upon by multiple parties and can provide a cornerstone of an interoperable distributed IoP. Regardless of the phrasing, facts do not have to be true or valid in the real world (e.g., simulation results). Hence, this definition only applies to the data itself without semantically checking the content.

In practice, however, data is rarely immutable. In fact, most data changes over time due to various reasons and even traditionally very static content, such as academic papers, are regularly published in several different versions over time (e.g., [45]). As such, considering all data to be a series of data revisions at different points in time, i.e., time series data, is a more reasonable approach. An analysis of successful existing collaborative, distributed information systems for tasks such as distributed version control [46]), software package management [47], [48] or research data management [49], [50], reveals that creating immutable, referenceable and uniquely identifiable data revisions (i.e., facts) is indeed a shared pattern of these systems.

D. Lifting Data to Facts

Thus, to enable reliable collaboration, interoperability and reuse in an IoP, we propose the following procedure to 'lift' any arbitrary data point to a fact:

1) Ensure the data has an (internal) unique identifier $\texttt{id}$, by which it is (internally) retrievable.
2) Determine the fixed identifier $\texttt{auth}$ of the authoritative source of the data.
3) Convert the data to an immutable fact by using a suitable revisioning system to assign each version of the data a revision identifier $\texttt{rev}$.

By implementing this approach, we further address requirement **RQ3** and can directly employ revision information in the fact identifiers to improve the reliability of information.

E. FactID: Persistent Identification of Facts

The triple of *authority ID*, *internal ID* and *revision ID* $\langle\texttt{auth},\texttt{id},\texttt{rev}\rangle$ serves as a *persistent identifier (PID)* for an immutable fact $f_{\langle\texttt{auth},\texttt{id},\texttt{rev}\rangle}$ and indefinitely references the exact same data. As such, it will still be a valid persistent identifier even if the identified fact is no longer available or was actively deleted. We refer to this (unique) persistent identifier of a fact as its **FactID**.

An authority (identified by $\langle\texttt{auth}\rangle$) can be responsible for any number of data objects (identified through the pair $\langle\texttt{auth},\texttt{id}\rangle$), which in turn can have any number of revisions $\langle\texttt{auth},\texttt{id},\texttt{rev}\rangle$. Due to their immutability, facts are perfectly suitable for caching and archiving (e.g., using external data archiving services).

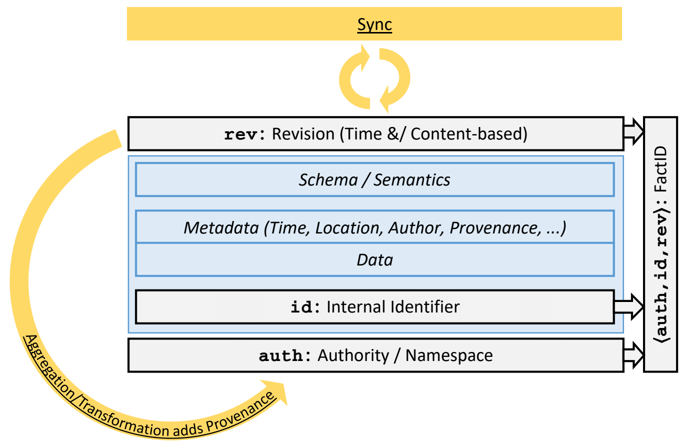


Fig. 1. Basic layer model of facts as a triple of (i) authority, (ii) (meta) data, and (iii) revision, enabling persistent referencing and master/slave synchronization based on sequence of immutable data revisions, as well as provenance tracking via references to source facts.

In the simplest case, the revision identifier ($\langle\texttt{rev}\rangle$) is the timestamp when this revision of the data object was created, but implementing other – e.g., content-based – identifiers such as hashes etc., is acceptable, as long as a way to determine the 'most current' of any two revisions of the same data (as identified by the authority and internal data identifier) exists. More formally, a given data object, identified by its $\langle\texttt{auth},\texttt{id}\rangle$ identifier pair, and the set of all associated facts (i.e., corresponding revisions) $F_{\langle\texttt{auth},\texttt{id}\rangle}$, there must exist a total order, i.e., a binary relation O on these revisions, such that for all $a, b, c \in F_{\langle\texttt{auth},\texttt{id}\rangle}$:

$$\begin{array}{ccccc} O(a,b) & \wedge & O(b,a) & \Rightarrow & a=b \qquad (antisymmetry) \\ O(a,b) & \wedge & O(b,c) & \Rightarrow & O(a,c) \qquad (transitivity) \\ O(a,b) & \vee & O(b,a) & \Leftrightarrow & true \qquad (connexity) \end{array}$$

Since revisions provide a natural ordering, determining the availability of newer versions is trivial, as well as discovering the latest revision of any given data point. As such master/slave replication and synchronization systems can be built on top of the revisioning primitive with minimal effort.

Finally, facts can be semantically enriched using domain-specific data models and ontologies, where applicable, to improve interoperability and machine interpretability, as well as to enable the transfer of semantic context information. We refer to such semantically enriched facts as **semantic facts** (addressing **I8**).

As illustrated in Figure 1, this definition leads to a basic layer model where a fact is an individual **revision** of data (consisting of an identifier, the actual data and meta data and optional schema and semantic information) under the domain of an authority. Since the facts are directly associated with corresponding PIDs, they can be referenced and resolved across domains and authorities, linked, aggregated, and reused across system and organizational boundaries.

F. FactDAG: Tracking Fact Provenance over Time

Adopting the FAIR principle R1.2 and thus, addressing issue **I7**, requires to register the PIDs of the source facts in the provenance information whenever data is combined, aggregated, or transformed. This tracking of facts' provenance information directly results in the creation of a directed acyclic graph (DAG) of facts, or simply a **FactDAG**. Such a graph of provenance information allows for better judgement of the formation history, the credibility of facts, as well as supporting to trace back origins of phenomena in data, similar to how e.g., the Git commit history works [46].

Formally, a FactDAG employing only a single type of influence edge to track provenance can be defined as a set of facts F and corresponding directed influence edges $I \subset F \times F$, such that it is cycle-free, i.e., that there is no sequence of directed influence edges $(f_1, f_2), (f_2, f_3), ..., (f_n, f_1) \in I$ such that whenever fact f_i is in this sequence for all $1 \leq i < n$, then there is an influence edge (f_i, f_{i+1}).

Together with information about the actor that inflicted a change, as well as the processes involved, an **audit log** of all relevant changes is constructed. A viable candidate for the implementation of such a log is the established W3C standard PROV [51]. Hereby, processes and actors may themselves simply be facts, which are specifically referenced in the provenance information of the derived fact.

To address the requirement **RQ4**, a suitable approach is to rely on usage policy languages, such the SPECIAL Usage Policy Language [52] or CPPL [53], which additionally allows for the modeling and tracking of data usage and usage consent. Implementations employing distributed ledgers, such as the IOTA Tangle [54], for tamper resistance and decentralized trust seem equally feasible. Regardless, of the particular implementation, a single approach should be pursued to enable interoperability on a large scale. Unfortunately, these aspects are not legally binding yet and therefore, they mandate legislative changes to force data processors to respect data sovereignty wishes and to allow for fines in case of violations. The recent introduction of the GDPR [55] shows that related legislative approaches can be successful.

Furthermore, the issue **I3** concerning data reliability can be tackled through similar means. While attached meta information can classify the quality of information, legislative boundaries can make sure that only usable information is being shared with collaborators [43]. For example, the expiration date of products already forbids the sale of old physical goods to protect consumers. Such an approach could be transferred to the digital domain of production data as well to shield users from inaccurate information. Furthermore, digital signatures may be used to technically enforce an authorities' responsibility for their data and ensure its authenticity, as well as cryptography to prevent unauthorized access [56]. However, since the FactDAG model is a conceptual data layer model, transport and other security concerns must be addressed by future implementations or left to the transport layer. Consequentially, our design of a formalized FactDAG model is functional wrt. data interoperability regardless of any previously agreed upon, individual security implementation.

IV. APPLICATION

The application of the basic layer model of facts sets a standard in data provisioning whereas the *FactDAG* enables the interorganizational exchange of provenance-linked facts, and therefore the reliable and accountable adumbration of manufacturing processes.

For production environments that actively apply data-driven models to support the decision making and optimize machines or material usage, a transparent and auditable decision making is important. In this context, facts represent reliable data (even when it originates from external sources) and the *FactDAG* illustrates the flow of sensory information's transformations up to the point where the information is used for a decision.

In an automation process, facts describe control information, e.g., the state of a machine or certain information about a product while the *FactDAG* describes the interaction, alteration and dependencies of these facts. As a result, analyzing a *FactDAG* consisting of control information and commands, enables the analysis of the dependencies in a flow of control information over time.

A. Exposing Information as Facts

By representing different revisions of data as a series of facts, collaborative interorganizational reuse of data can be facilitated. To enable interoperability with existing information systems, lifting data that stores information to facts can be accomplished by employing gateway services, caching and versioning existing resources on-the-fly using a mechanism similar to the HTTP Memento protocol [57], [58]. Such an approach could also enable the usage of third-party archiving services and backwards compatibility with data sources not explicitly implementing revisioning and the FactDAG model, i.e., data can be transformed before being shared in an interorganizational setting. Hence, no changes to today's deployed sensors and devices that record data are required to support a sharing of facts.

B. Composing FactIDs from Existing Identifiers

In practice, the abstract notions of *authority ID*, *internal ID* and *revision ID* must be implemented in terms of concrete identifier systems, favorably compatible with existing infrastructure. As such, a conceivable practical implementation of a *FactID* could reuse domain names [59] as authority ID, URI paths [60] or other existing object identifiers as internal ID and timestamps as revision ID. Hence, the introduction of FactIDs does not have to go along with an introduction of new labeling approaches.

An example of such a FactID could consist of the domain name `wzl.rwth.de`, the path `machineX/partY/sensorZ`, and the timestamp *2019/07/11 17:07:27*, e.g., represented as UNIX timestamp `1562857647`. Thus, constructing the FactID ⟨`wzl.rwth.de,machineX/partY/sensorZ,1562857647`⟩. Subsequently, FactIDs can be created in a way that they are compatible with existing production systems deployments and information systems, while they are reusing existing and easily (human-)interpretable identifiers.

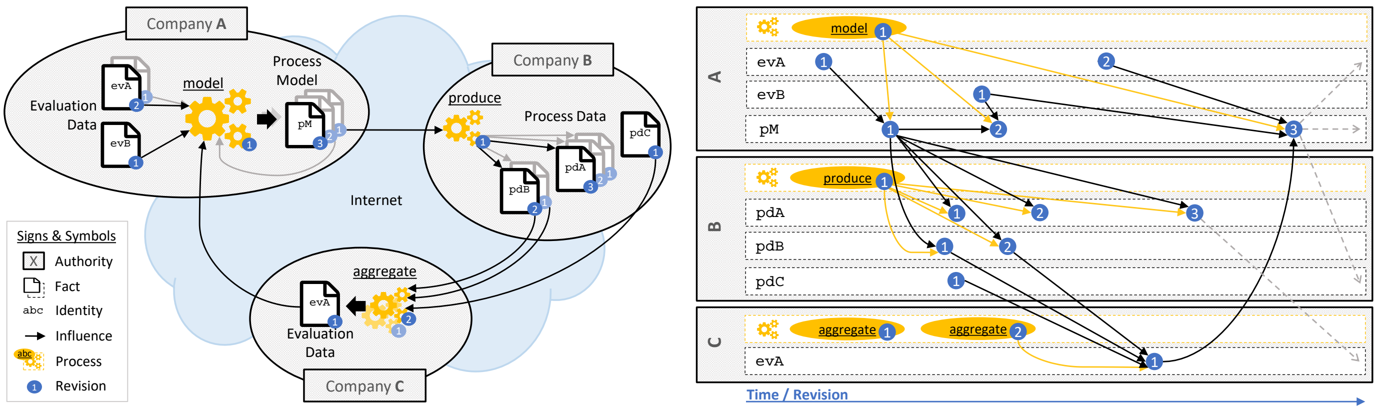


Fig. 2. An exemplary business process over time involving three Companies **A**, **B**, and **C**, collaborating in a WWL (illustrated on the left), and a corresponding *FactDAG* (visualized on the right) that highlights the evolution of data and process revisions as well as their influences. Each information is uniquely identified by its corresponding FactID. These facts can in turn also be utilized further (dashed arrows).

C. Tracking Dataflows in the FactDAG

By incorporating facts representing the activities or processes leading to data changes into the FactDAG, it is possible to track and capture the physical and logical dataflows in the FactDAG.

In Figure 2, we illustrate how a FactDAG and dataflows relate in an exemplary business process. In the presented scenario, company $\langle\mathtt{A}\rangle$ continuously designs, evaluates, and models $\langle\mathtt{A},\mathtt{model}\rangle$ production processes, deriving an optimized process model $\langle\mathtt{A},\mathtt{pM}\rangle$ from (both internally and externally acquired) evaluation data $\langle\mathtt{A},\mathtt{evA}\rangle$, $\langle\mathtt{A},\mathtt{evB}\rangle$ and $\langle\mathtt{C},\mathtt{evA}\rangle$. Under its authority, the initial revision of the model $\langle\mathtt{A},\mathtt{pM},1\rangle$ is derived exclusively from fact $\langle\mathtt{A},\mathtt{evA},1\rangle$, but subsequently incrementally updated by also including $\langle\mathtt{A},\mathtt{evB},1\rangle$ as that fact becomes available, resulting in $\langle\mathtt{A},\mathtt{pM},2\rangle$. Later a new revision $\langle\mathtt{A},\mathtt{evA},2\rangle$ of $\langle\mathtt{A},\mathtt{evA}\rangle$ is created and additional evaluation data $\langle\mathtt{C},\mathtt{evA}\rangle$ is externally acquired from company $\langle\mathtt{C}\rangle$. Finally, a new model revision $\langle\mathtt{A},\mathtt{pM},3\rangle$ is derived from scratch, solely based on facts $\langle\mathtt{A},\mathtt{evA},2\rangle$, $\langle\mathtt{A},\mathtt{evB},1\rangle$, and $\langle\mathtt{C},\mathtt{evA},1\rangle$. All revisions of model $\langle\mathtt{A},\mathtt{pM}\rangle$ are derived using the initial revision of process $\langle\mathtt{A},\mathtt{model},1\rangle$.

Additionally, company $\langle\mathtt{B}\rangle$ produces $\langle\mathtt{B},\mathtt{produce}\rangle$ goods using (the initial revision of) company $\langle\mathtt{A}\rangle$'s process model $\langle\mathtt{A},\mathtt{pM},1\rangle$, simultaneously producing process data $\langle\mathtt{B},\mathtt{pdA}\rangle$ and $\langle\mathtt{B},\mathtt{pdB}\rangle$. Company $\langle\mathtt{B}\rangle$ further provides process data from similar related processes $\langle\mathtt{B},\mathtt{pdC}\rangle$. Under its authority, three revisions of process data $\langle\mathtt{B},\mathtt{pdA}\rangle$ and two revisions of process data $\langle\mathtt{B},\mathtt{pdB}\rangle$ are derived employing process $\langle\mathtt{B},\mathtt{produce},1\rangle$, as well as an additional data point $\langle\mathtt{B},\mathtt{pdC}\rangle$ without explicit information of the involved processes.

Company $\langle\mathtt{C}\rangle$ is a service provider that aggregates $\langle\mathtt{C},\mathtt{aggregate}\rangle$ process data from various sources (here: $\langle\mathtt{B},\mathtt{pdB}\rangle$ and $\langle\mathtt{B},\mathtt{pdC}\rangle$ from company $\langle\mathtt{B}\rangle$) and derives additional evaluation data $\langle\mathtt{C},\mathtt{evA}\rangle$ that it sells to its customers. Hereby, company $\langle\mathtt{C}\rangle$ explicitly makes use of the possibility to interpret multiple revisions of $\langle\mathtt{B},\mathtt{pdB}\rangle$ as a time series of data through unique FactIDs.

Overall, our exemplary presentation only provides a partial view on all (global) dependencies. To indicate further dataflows, we include additional exemplary dashed gray arrows in Figure 2. While the real-world business process includes dataflow cycles (i.e., data flowing back to a data source, tightly incorporated into the business process), the *FactDAG* remains cycle-free with edges only pointing in one direction along the time-axis. As such, the integrity of references, provenance information and attributions remains intact over time while cyclic dataflows can be modeled more reliably than using traditional ETL pipelines, i.e., satisfying the needs of data interoperability in an Internet of Production.

D. FactDAG Applications in Manufacturing

To illustrate the advantages of a *FactDAG* (as sketched in Section III-F), we map the abstract schema to a concrete application based on real-world companies in different exemplary use cases. Suppose that we operate in a scenario with a steel manufacturer $\langle\mathtt{Arcalor}\rangle$, a manufacturer of components $\langle\mathtt{Busch}\rangle$, and an automotive company using these components in an automated assembly line $\langle\mathtt{CMW}\rangle$. Manufacturer $\langle\mathtt{Arcalor}\rangle$ produces blocks of steel with a certain tolerance regarding its quality features. During the manufacturing process, the manufacturer gathers sensor data, e.g., rolling forces $\langle\mathtt{Arcalor},\mathtt{roll/F}\rangle$, temperature gradients $\langle\mathtt{Arcalor},\mathtt{roll/dTemp}\rangle$, or acoustic emissions. The manufacturer also employs physical models and simulations $\langle\mathtt{Arcalor},\mathtt{roll/model}\rangle$ that describe the behaviour of the material during the manufacturing processes. After a block of steel has been produced, all gathered data is used to predict certain hardness conditions of the material at different positions, i.e., $\langle\mathtt{Arcalor},\mathtt{workpiece/99/hardness},1\rangle$ for the workpiece with internal work order number `99` and revision ID `1`.

Component manufacturer $\langle\mathtt{Busch}\rangle$, who processes that block of steel for the manufacturing of a component $\langle\mathtt{Busch},\mathtt{component/42}\rangle$, can then utilize the information $\langle\mathtt{Arcalor},\mathtt{workpiece/99/hardness},1\rangle$ provided by the material manufacturer $\langle\mathtt{Arcalor}\rangle$ to adapt its process to each specific workpiece by adjusting the process forces or the path taken during, e.g., a milling process, to prevent, e.g., tool deflection (cf. Section II-B2). At the same time, during the milling process, the manufacturer gathers geometric properties $\langle\mathtt{Busch},\mathtt{component/42/geometry},1\rangle$ of the workpiece as well as other relevant process parameters ⟨`Busch, component/42/parameters, 1`⟩.

Company ⟨`CMW`⟩ uses that data to automatically adjust its adaptive assembly line to the workpiece dimensions and requirements. Here, sensor systems acquire data that is processed to control movements or operations of robots to assemble different components together.

Suppose company ⟨`CMW`⟩ now detects, that component ⟨`Busch, component/42`⟩ fails to pass internal quality control. In order to determine all processes and parties involved in the creation of the component, the recorded provenance directly allows backtracing all potentially relevant influences across organizational boundaries as the closure over the influence edges that lead to the component's creation. This formation history of the flawed component may further be directly compared to the formation history of similar, e.g., unflawed workpieces in order to detect deviations which may indicate a potential source of error. Suppose that a specific revision `X` of the temperature gradients ⟨`Arcalor, roll/dTemp`⟩ describing the steel which was used to manufacture component ⟨`Busch, component/42`⟩ reveals, that the steel used to create the faulty component was quenched abnormally fast, resulting in different material characteristics just within the tolerance specified by component manufacturer ⟨`Busch`⟩ but insufficient for the intended application at company ⟨`CMW`⟩. Subsequently ⟨`CMW`⟩ may e.g., either (directly or indirectly through ⟨`Busch`⟩) provide additional manufacturing parameters ⟨`CMW, requirements`⟩ to ⟨`Arcalor`⟩ or automatically adapt its internal assembly process to incorporate additional preprocessing and preparation steps to ensure the required quality parameters.

As such, process data in the FactDAG provide valuable context information and provenance for customers and producers alike, by providing a conceptual foundation for the bidirectional information flow between suppliers and customers across interorganizational boundaries. The establishment of cyclic information flow is also a key enabler for agile manufacturing [61], i.e., the iterative product evaluation and improvement within the user cycle and based on usage data (cf. Section II-B1).

Without a structured and context-aware representation of information, such improvements are unlikely to be realized [7], especially when interorganizational collaboration is not already deeply integrated into the business process [8].

The above example illustrates advantages for automated assembly lines, fine blanking lines as well as machine tool optimizations in general given an implemented FactDAG for data interoperability. Every manufacturer of a semi-finished product can evaluate the performance of its product by accessing upstream information connected to its product, whereas the user of the semi-finished product may adjust its process precisely to the individual properties of the delivered product. Thus, if each company would share their information, every manufacturing process could be integrated into the concept of a WWL that would significantly increase the amount of available data and the opportunities for data-driven modeling of manufacturing systems. Additionally, such a system allows tracking audits of the origin, history of origin and version of each data used and generated in any system to ensure liability and trust in the data resulting from uncertainties due to interorganizational collaborations.

V. Conclusion

In the future, production will strongly rely on digital data as nowadays already exhibited by the Internet of Things. Hence, for each physical workpiece, a manufacturer will also provide a digital manifest. By establishing a Worldwide Lab, an Internet of Production enables (cross-domain) collaboration between different departments, production sites, or even between companies to improve their well-established processes. However, existing approaches fail to support data interoperability which would facilitate additional improvements of production processes. By describing different industrial use cases, we discussed the high variability of process data and derived the need for data interoperability. Overall, we identified *ten* key issues with today's data layer models.

We hypothesized that existing interoperability systems fundamentally suffer from mutability of data and the lack of clear data authorities. Hence, we derived a set of *five* requirements for (Internet-scale) data interoperability, extending the well-known FAIR data principles. Subsequently, we introduced a data interoperability model design, addressing the identified key issues. The deep incorporation of provenance information into this model enables companies to rely on persistent identifiability, synchronization, data sovereignty, and accountability, even in interorganizational scenarios. We refer to this model as *FactDAG*. To illustrate the impact of our presented approach, we highlighted (immediate) advantages of its implementation in a Worldwide Lab.

In sight of large prospective advantages of the proposed approach for an Internet of Production, future work will focus on the design of a concrete implementation of the model, carefully considering the requirements and circumstances of the abundance of different existing production systems, communication protocols, interaction patterns and data formats encountered in production environments, as well as the evaluation of its practical effectiveness. As such, the *FactDAG* model serves as ground work for other applications in production systems and therefore will be used as an enabler for the concepts to make data findable, accessible, interoperable and reusable in an Internet of Production.

References

[1] S. Jeschke, C. Brecher, T. Meisen, D. Özdemir, and T. Eschert, "Industrial Internet of Things and Cyber Manufacturing Systems," in *Industrial Internet of Things*, 2017.

[2] J. Lee, B. Bagheri, and H.-A. Kao, "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems," *Manuf. Lett.*, vol. 3, 2015.

[3] R. Baheti and H. Gill, "Cyber-physical Systems," *IEEE The Impact of Control Technology*, vol. 12, no. 1, 2011.

[4] P. Heimes, S. Senel *et al.*, "IoT Platforms for the Internet of Production," *IEEE IoT-J*, 2018.

[5] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of Big data on cloud computing: Review and open research issues," *Information Systems*, vol. 47, 2015.

[6] R. Glebke, M. Henze, K. Wehrle, P. Niemietz, D. Trauth, P. Mattfeld, and T. Bergs, "A Case for Integrated Data Processing in Large-Scale Cyber-Physical Systems," in *HICSS*, 2019.

[7] R. Anderl, "Industrie 4.0 – Advanced Engineering of Smart Products and Smart Production," in *19th Seminar on High Technology*, 2014.

[8] J. M. Bryson, B. C. Crosby, and M. M. Stone, "Designing and Implementing Cross-Sector Collaborations: Needed and Challenging," *Public Administration Review*, vol. 75, no. 5, 2015.

[9] H. Singh, R. K. Garg, and A. Sachdeva, "Supply chain collaboration: A state-of-the-art literature review," *Uncertain Supply Chain Management*, vol. 6, no. 2, 2018.

[10] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, Opportunities, and Directions," *IEEE Trans. Industr. Inform.*, vol. 14, no. 11, 2018.

[11] J. Pennekamp, M. Henze, S. Schmidt, P. Niemietz, M. Fey, D. Trauth, T. Bergs, C. Brecher, and K. Wehrle, "Dataflow Challenges in an Internet of Production: A Security & Privacy Perspective," in *ACM CPS-SPC*, 2019.

[12] J. Pennekamp, R. Glebke, M. Henze, T. Meisen, C. Quix, R. Hai, L. Gleim, P. Niemietz, M. Rudack, S. Knape, A. Epple, D. Trauth, U. Vroomen, T. Bergs, C. Brecher, A. Bührig-Polaczek, M. Jarke, and K. Wehrle, "Towards an Infrastructure Enabling the Internet of Production," in *IEEE ICPS*, 2019.

[13] A. Bosche, D. Crawford, D. Jackson, M. Schallehn, and P. Smith, "How Providers Can Succeed in the Internet of Things," Bain & Company, 2016.

[14] P. Mell and T. Grance, "The NIST definition of cloud computing," *National Institute of Standards and Technology*, vol. 53, no. 6, 2009.

[15] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, 2010.

[16] A. Bröring, S. Schmid, C.-K. Schindhelm, A. Khelil, S. Kabisch, D. Kramer, D. Le Phuoc, J. Mitic, D. Anicic, and E. Teniente López, "Enabling IoT Ecosystems through Platform Interoperability," *IEEE Software*, vol. 34, no. 1, 2017.

[17] O. Givehchi, K. Landsdorf, P. Simoens, and A. W. Colombo, "Interoperability for Industrial Cyber-Physical Systems: An Approach for Legacy Systems," *IEEE Trans. Industr. Inform.*, vol. 13, no. 6, 2017.

[18] P. Glanon, S. Azaiez, and C. Mraidha, "A Modular Interoperability Layer for Connecting the Business and Manufacturing Systems," in *IEEE WFCS*, 2018.

[19] M. Schneider, B. Hippchen, S. Abeck, M. Jacoby, and R. Herzog, "Enabling IoT Platform Interoperability Using a Systematic Development Approach by Example," in *GIoTS*, 2018.

[20] J. Nilsson and F. Sandin, "Semantic Interoperability in Industry 4.0: Survey of Recent Developments and Outlook," in *IEEE INDIN*, 2018.

[21] G. Pardo-Castellote, "OMG Data-Distribution Service: Architectural Overview," in *ICDCS*, 2003.

[22] A. Vijayaraghavan, W. Sobel, A. Fox, D. Dornfeld, and P. Warndorf, "Improving Machine Tool Interoperability Using Standardized Interface Protocols: MT Connect," *Green Manufacturing and Sustainable Manufacturing Partnership*, 2008.

[23] W. Mahnke, S.-H. Leitner, and M. Damm, *OPC Unified Architecture*. Springer Science & Business Media, 2009.

[24] B. Otto, S. Auer, J. Cirullies, J. Jürjens, N. Menz, J. Schon, and S. Wenzel, "Industrial Data Space: Digital Souvereignity over Data," *Fraunhofer White Paper*, 2016.

[25] D. Gürdür, F. Asplund, and J. El-khoury, "Measuring tool chain interoperability in Cyber-Physical Systems," in *SoSE*, 2016.

[26] N. C. Shu, B. C. Housel, R. W. Taylor, S. P. Ghosh, and V. Y. Lum, "EXPRESS: a data EXtraction, Processing, and Restructuring System," *ACM TODS*, vol. 2, no. 2, 1977.

[27] P. Vassiliadis, "A Survey of Extract–Transform–Load Technology," *IJDWM*, vol. 5, no. 3, 2009.

[28] F. Klocke and A. Kuchle, *Manufacturing Processes*. Springer, 2009.

[29] Q. Zheng, X. Zhuang, and Z. Zhao, "State-of-the-art and future challenge in fine-blanking technology," *Production Engineering*, 2018.

[30] H. Voigts, D. Trauth, A. Feuerhack, P. Mattfeld, and F. Klocke, "Dependencies of the die-roll height during fine blanking of case hardening steel 16MnCr5 without V-ring using a nesting strategy," *Int. J. Adv. Manuf. Technol.*, vol. 95, no. 5, 2018.

[31] D. Harsch, J. Heingärtner, Y. Renkci, and P. Hora, "Influence of scattering material properties on the robustness of deep drawing processes," in *FTF*, 2017.

[32] T. Bergs, P. Niemietz, J. Pennekamp, I. Kunze, D. Trauth, and K. Wehrle, "Stamping Process Modelling in an Internet of Production," in *CIRP TESConf*, 2019.

[33] M. Königs and C. Brecher, "Process-parallel virtual quality evaluation for metal cutting in series production," *Procedia Manuf.*, vol. 26, 2018.

[34] C. Brecher, M. Buchsbaum, and S. Storms, "Control from the Cloud: Edge Computing, Services and Digital Shadow for Automation Technologies," in *ICRA*, 2019.

[35] J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler, "Named Graphs, Provenance and Trust," in *WWW*, 2005.

[36] W.-T. Tsai, X. Wei, Y. Chen, R. Paul, J.-Y. Chung, and D. Zhang, "Data provenance in SOA: security, reliability, and integrity," *Service Oriented Computing and Applications*, vol. 1, no. 4, 2007.

[37] M. Conti, S. K. Das, C. Bisdikian, M. Kumar, L. M. Ni, A. Passarella, G. Roussos, G. Tröster, G. Tsudik, and F. Zambonelli, "Looking ahead in pervasive computing: Challenges and opportunities in the era of cyberphysical convergence," *Pervasive and Mobile Computing*, vol. 8, no. 1, 2012.

[38] J. C. Wallis, E. Rolando, and C. L. Borgman, "If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology," *PloS one*, vol. 8, no. 7, 2013.

[39] S. Bechhofer, I. Buchan, D. De Roure, P. Missier, J. Ainsworth, J. Bhagat, P. Couch, D. Cruickshank, M. Delderfield, I. Dunlop *et al.*, "Why Linked Data is Not Enough for Scientists," *Future Generation Computer Systems*, vol. 29, no. 2, 2013.

[40] L. L. Haak, "Persistent identifiers can improve provenance and attribution and encourage sharing of research results," *Inf. Serv. Use*, vol. 34, no. 1-2, 2014.

[41] R. Mordinyi and S. Biffl, "Versioning in Cyber-physical Production System Engineering – Best-Practice and Research Agenda," in *IEEE/ACM SEsCPS*, 2015.

[42] J. Fitzgerald, C. Gamble, P. G. Larsen, K. Pierce, and J. Woodcock, "Cyber-Physical Systems Design: Formal Foundations, Methods and Integrated Tool Chains," in *IEEE/ACM FormaliSE*, 2015.

[43] J. Pennekamp, M. Dahlmanns, L. Gleim, S. Decker, and K. Wehrle, "Security Considerations for Collaborations in an Industrial IoT-based Lab of Labs," in *IEEE GCIoT*, 2019.

[44] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, 2016.

[45] L. H. Nielsen, "Zenodo now supports DOI versioning!" 2017. [Online]. Available: http://web.archive.org/web/20190207141003/http://blog.zenodo.org/2017/05/30/doi-versioning-launched/

[46] J. Loeliger and M. McCullough, *Version Control with Git: Powerful tools and techniques for collaborative software development*. O'Reilly Media, 2012.

[47] T. Gamblin, M. LeGendre, M. R. Collette, G. L. Lee, A. Moody, B. R. de Supinski, and S. Futral, "The Spack Package Manager: Bringing Order to HPC Software Chaos," in *ACM SC*, 2015.

[48] J. A. Forbes, J. D. Stone, S. Parthasarathy, M. J. Toutonghi, and M. V. Sliger, "Software package management," 2002, uS Patent 6,381,742.

[49] P. Herterich and S. Dallmeier-Tiessen, "Data Citation Services in the High-Energy Physics Community," *D-Lib Magazine*, vol. 22, no. 1/2, 2016.

[50] "Zenodo - Research. Shared." 2018. [Online]. Available: http://web.archive.org/web/20181204082805/http://help.zenodo.org/#versioning

[51] P. Groth and L. Moreau, "PROV-Overview," *W3C Working Group Note 30 April 2013*, 2013.

[52] S. Kirrane, J. D. Fernández, W. Dullaert, U. Milosevic, A. Polleres, P. A. Bonatti, R. Wenning, O. Drozd, and P. Raschke, "A Scalable Consent, Transparency and Compliance Architecture," in *ESWC*, 2018.

[53] M. Henze, J. Hiller, S. Schmerling, J. H. Ziegeldorf, and K. Wehrle, "CPPL: Compact Privacy Policy Language," in *WPES*, 2016.

[54] S. Popov, "The Tangle," *White Paper*, 2016.

[55] P. Voigt and A. Von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*, 1st ed. Springer, 2017.

[56] C. F. Kerry and C. Romine, "FIPS PUB 186-4 Federal Information Processing Standards Publication: Digital Signature Standard (DSS)," 2013.

[57] H. Van de Sompel, M. L. Nelson, R. Sanderson, L. L. Balakireva, S. Ainsworth, and H. Shankar, "Memento: Time Travel for the Web," arXiv preprint arXiv:0911.1112 [cs.CR], 2009.

[58] H. Van de Sompel, M. L. Nelson, and R. Sanderson, "HTTP Framework for Time-Based Access to Resource States – Memento," RFC 7089, 2013.

[59] P. Mockapetris and K. J. Dunlap, "Development of the Domain Name System," in *ACM SIGCOMM*, 1988.

[60] T. Berners-Lee, R. T. Fielding, and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax," RFC 3986, 2005.

[61] M. Brettel, M. Klein, and N. Friederichsen, "The relevance of manufacturing flexibility in the context of Industrie 4.0," *Procedia CIRP*, vol. 41, 2016.

Lars Gleim received the B.Sc. and M.Sc. degrees in Computer Science from RWTH Aachen University. He is a researcher at the Chair of Databases and Information Systems (DBIS) at RWTH Aachen University. His research focuses on semantic information systems for interorganizational interoperability in the context of the Industrial Internet of Things (IIoT). His special interests include applications of statistical methods and web technologies in order to lower entry barriers of semantic interoperability.

Jan Pennekamp received the B.Sc. and M.Sc. degrees in Computer Science from RWTH Aachen University with honors. He is a researcher at the Chair of Communication and Distributed Systems (COMSYS) at RWTH Aachen University. His research focuses on security & privacy aspects in the Industrial Internet of Things (IIoT). In particular, his special interests include privacy-enhancing technologies, the design of privacy-preserving protocols, and secure computations as well as their application.

Martin Liebenberg received the Diploma (equiv. M.Sc.) degree in Computer Science from RWTH Aachen University. He is a researcher in the Knowledge-Based Systems Group (KBSG) at the Chair of Databases and Information Systems at RWTH Aachen University. His research focuses on applied artificial intelligence in the context of industrial production. In particular, his special interests include optimization of production processes and decision support.

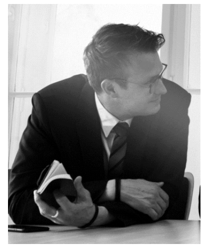
Melanie Buchsbaum received the B.Sc. degree in Mechanical Engineering and M.Sc. degree in Automation Engineering from RWTH Aachen University. She is a researcher at the Chair of Machine Tools at the Laboratory for Machine Tools and Production Engineering (WZL) at RWTH Aachen University. Her research focuses on automation in the Industrial Internet of Things (IIoT). In particular, her special interests include Cloud and Edge Computing as enabler for the Digital Shadow.

Philipp Niemietz received the B.Sc. and M.Sc. degrees in Computer Science from RWTH Aachen University with honors. He is research group leader at the Chair for Manufacturing Technologies at the Laboratory for Machine Tools and Production Engineering WZL at RWTH Aachen University. His research focuses on data driven modelling of cyber physical systems in context of manufacturing processes, in particular for stamping processes.

Simon Knape received the B.Sc. degree in Mechanical Engineering and M.Sc. degree in Production Engineering from RWTH Aachen University. He has been a researcher at the Chair of Machine Tools at the Laboratory for Machine Tools and Production Engineering (WZL) at RWTH Aachen University. His research focuses on machine data analytics for machine tools. In particular, his special interests include Edge Computing and database systems as enabler for the Digital Shadow.

Alexander Epple studied Mechanical Engineering at RWTH Aachen University. Since 2012 he works with the chair for machine tools at the Laboratory for Machine Tools and Production Engineering (WZL) at the RWTH. In 2014 he became head of the research group "Design and Simulation of Manufacturing Systems". Since 2016 he has been chief engineer at the WZL for Machine Data Analytics and NC-Technology and his doctorate was awarded in November 2017.

Simon Storms acquired the M.Sc. in Automation Engineering at the faculty for Mechanical Engineering at the RWTH Aachen. From 2014 to 2017 he was a researcher at the chair for machine tools at the Laboratory for Machine Tools and Production Engineering (WZL) at the RWTH Aachen University focusing robotics and automation. In 2017 he became head of the research group "Automation". Since 2016 he has been chief engineer at the WZL for Automation and Control.



Daniel Trauth studied Mechanical Engineering and Business Economics at the RWTH Aachen University, where he also received his PhD in machine hammer peening and tribology. Since 2017 he has been chief engineer for Grinding, Forming and Planning Technology at WZL, head of the working group machine learning in production engineering and since 2019 leads WZL's Digital Transformation in Manufacturing Technologies department.

Thomas Bergs studied Mechanical Engineering at RWTH Aachen University. In 2001 he received his PhD on laser-assisted machining, working at the Fraunhofer Institute for Production Technology IPT. He is full professor of Mechanical Engineering at the Chair of Manufacturing Technologies, RWTH Aachen University and board member of IPT and WZL at RWTH. His research focuses on fundamentals of manufacturing technology, process analyses, monitoring, modelling and simulation.

Christian Brecher received his doctorate from the Machine Technology Department at the Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University in 2001. He is full professor of Mechanical Engineering at RWTH Aachen University, Member of the Board of Directors of WZL and acting Head of Institute at IPT (Fraunhofer Institute for Production Technology). His areas of specialist expertise include machine, transmission and control systems technology.

Stefan Decker received his doctorate from Karlsruhe Technical University. He is a full professor heading the chair of Databases and Information Systems at RWTH Aachen University and Director of the Fraunhofer Institute for Applied Information Technology FIT. His research interests include the Semantic Web, metadata, ontologies and semi-structured data, web services, and applications for Digital Libraries, Knowledge Management, Information Integration and Peer-to-Peer technology.

Gerhard Lakemeyer received his PhD in Computer Science from the University of Toronto. He is a Professor leading the Knowledge-Based Systems Group in the Department of Computer Science at RWTH Aachen University. His special research interests include Artifical Intelligence, Knowledge Representation and Cognitive Robotics.

Klaus Wehrle received the Diploma (equiv. M.Sc.) and PhD degrees from University of Karlsruhe (now KIT), both with honors. He is full professor of Computer Science and Head of the Chair of Communication and Distributed Systems (COMSYS) at RWTH Aachen University. His research interests include (but are not limited to) engineering of networking protocols, (formal) methods for protocol engineering and network analysis, reliable communication software, and all operating system issues of networking.