DESTINI

SMART DATA PROCESSING AND SYSTEMS OF DEEP INSIGHT

http://www.destini2020.eu

Deliverable D2.1

Survey on Smart Data Processing and Systems of Deep Insight:

Current Research and Future Challenges

Document details:

Editor:	Andreas Andreou
Contributors:	Andreas Andreou, Michael Pingos, Panayiotis Christodoulou, Spyros Loizou, Andreas Christophorou
Date:	March 20, 2020
Version:	9.0

Document history:

Version	Date	Contributor	Comments
v.01	20/01/2020	Michalis Pingos, Spyros Loizou, Andreas Christophorou	Initial document structure
v.02	01/02/2020	Andreas Andreou	Review of document structure
v.03	10/02/2020	Panayiotis Christodoulou	First Draft
v.04	15/02/2020	Andreas Andreou	Comments
v.05	20/02/2020	Michalis Pingos	Corrections
v.06	25/02/2020	Panayiotis Christodoulou, Andreas Andreou	Final Comments
v.07	28/02/2020	Michalis Pingos	Final Corrections
v.08	02/03/2020	Partners	Review of document
v.09	20/03/2020	Panayiotis Christodoulou, Andreas Christophorou, Michalis Pingos	Approved final version after minor revision

Contents

List of figures

1. Introduction

Big Data (M. Chen, Mao, and Liu 2014) is an umbrella term referring to the large amounts of digital data continually generated by tools and machines, and the global population. The speed and frequency by which digital data is produced and collected by an increasing number of different kind of sources is projected to increase exponentially. This increasing volume of data, along with its immense social and economic value (Bertino 2013; Günther et al. 2017), is driving a global data revolution. Big Data have been called as “the new oil”, as it is recognized as a valuable human asset, which, with the proper collation and analysis, can deliver information that will give deep insights to many aspects of our everyday life and moreover to let us predict what might happen in the future.

While Big Data are available and easily accessible, it is evident that the great majority of them are coming from heterogeneous sources with irregular structure (Blazquez and Domenech 2018). Big Data originate mostly from one of the five primary sources: media, cloud, web, traditional business systems and Internet of Things (IoT) (Sethi and Sarangi 2017). Media includes social networks and interactive platforms, like Google, Twitter, Facebook, YouTube, Instagram, as well as generic media, like images, videos and audios that provide quantitative and qualitative data on every aspect of user interaction. Public or private cloud storages comprise information from real-time or on demand business data. Web or internet constitutes any type of data that are publicly available and can be used for any commercial or individual activity. Traditional business systems produce and store business data in conventional relational databases or modern NoSQL databases. Finally, IoT includes data generated from sensors that are connected to any electronic devices can emit data.

The process of transforming Big Data into Smart Data in terms of making them valuable and transforming them into meaningful information, is called Smart Data Processing (SDP) and includes a series of actions and techniques. These actions and techniques support the processing and integration of data into a unified view from the disparate Big Data sources. More specifically, the area of smart data processing includes the ability to clearly define, interoperate, openly share, access, transform, link, syndicate, and manage data. Under this perspective, it becomes crucial to have knowledge-based metadata representation techniques to structure the data sets and content, annotate them, link them with associated processes and software services, and deliver or syndicate information to recipients. This field also includes adaptive frameworks and tool-suites in support of smart data processing by allowing the best use of streaming or static data and may rely on advanced techniques for efficient resource management. The analytic solutions which rely on the smart data processing and integration techniques are called Systems of deep insight. These solutions enable optimization of asset performance in smart data processing systems and are geared towards systems of insight. In addition, they sift through the data to discover new relationships and patterns by analysing historical data, assessing current situation, applying business rules, predicting outcomes, and proposing the next best action.

This survey is conducted in the context of the DESTINI project and aims to identify and quote the most significant research findings, challenges and open problems on Smart Data Processing and Systems of Deep Insight approaches that are reported in the relevant literature. The rest of this document is structured as follows: Section 2 presents the research questions that motivated this study and describes the methodology followed to identify relevant studies published in various venues. Section 3 outlines the most important aspects of these studies organised in specific scientific areas, introducing the problem dealt, the methodology followed and the results produced in each of them. Section 4 summarizes the research challenges and open problems identified in the corresponding studies reviewed. Finally, section 5 concludes the survey.

2. Methodology and Materials

This section firstly presents the approach adopted to search and and classify relevant papers in literature that address a number of research questions posed in this study (see below) and secondly it outlines the main findings and results. Quantitative features are also provided in the form of bar charts to make it easier for the reader to identify the work performed categorised in areas, the venue as well as the year of publication.

2.1 Research Questions and Approach

This research study is motivated by two research questions:

RQ1: What is the scientific research devoted to the Smart Data Processing and Systems of Deep Insight area, with respect to the specific topics of the project's Joint Research Areas (JRAs)?

RQ2: Which are the main aspects of recent studies on the Smart Data Processing and Systems of Deep Insight area, with respect to the specific topics of the project's JRAs?

The three Joint Research Activities (JRAs) of the DESTINI's project are:

- Smart Data Processing Systems: This JRA includes data ingestion, data aggregation of an enormous variety of structured, unstructured and semi-structured datasets, knowledge-based meta-data representation techniques for the conversion of raw into smart data, data privacy and protection, automated deployment, run-time software performance monitoring and dynamic configuration.
- Systems of Deep Insight: This JRA focuses on analytic solutions that enable optimization of asset performance in smart data processing systems and is geared towards systems of insight. These are systems that turn data into insights, systematically test insights and find those data that matter to make them contextual and actionable.
- Methodology for Smart Data-centric Services & Applications: This JRA targets smart application development techniques by providing a methodology that interlocks elements of smart data processing and systems of deep insight to alleviate complexity and the effect of changes, thus speeding up the entire soft-ware development/deployment process for smart applications in priority sectors identified by the Smart Specialisation Strategy of Cyprus (S3Cy).

The topics per JRA of DESTINI are the following:

JRA1 - Smart Data Processing Systems

- Data ingestion and data curation
- Data aggregations

- Stream-processing
- Knowledge-based meta-data
- Scalable, reusable and secure processing framework

JRA2 - Systems of Deep Insight

- Large-scale data analytics
- Cross-correlation and cross functional models
- Actionable, context-dependent deep insights
- Decision monitoring and next best action
- Descriptive, predictive, cognitive analytics

JRA3 - Methodology for smart Data-centric Services & Applications

- Integrating smart data processing & and systems of deep insight in unified data-centric applications
- Agile methods and DevOps
- Smart Data-Centric Application and Data / Service Evolution

In order to provide answers to these research questions, a methodology was applied firstly to gather enough material and then assess the positions and statements made; this methodology is described in this section. As a first step, the guidelines for a systematic literature review (SLR) proposed by Kitchenham et al. (2010) was followed. Although a SLR is outside of the scope of this work, those guidelines assessed in organizing the process of finding and classifying relevant works. The search process aimed at locating articles indexed in Scopus, Science Direct, IEEE Xplore, ACM Digital Library, SpringerLink, Google Scholar and Wiley Online. The general search strings used were "Smart Data Processing Systems", "Systems of Deep Insight" and "Smart Data-centric Services and Applications". Additional, more refined searches were conducted using the following strings: "data ingestion", "data aggregation", "structured datasets", "unstructured datasets", "semi-structured datasets", "knowledge-based meta-data representation techniques", "conversion of raw data into smart data", "data privacy and protection", "run-time software performance monitoring and dynamic configuration", "big data", "data lakes ", "data warehouses ", "optimization in data processing", "data analytics", "business intelligence", "turn data into insights", "contextual and actionable data". The search results consisted of articles published up to 2019. As Smart Data Processing and Deep Insights is a very recent topic, both journal and conference articles were considered. Finally, duplicate papers were removed from the results, since the search engines and databases produced overlapping results to a certain extent. After these steps, the initial collection consisted of 81 potentially relevant works. Then, a detailed, qualitative analysis was performed by examining closely these papers in order to identify and merge different papers of the same authors/groups reporting their results incrementally and also works that used the term "Smart Data" with a different meaning compared to the target of this survey. In addition, the snowballing approach was used (see e.g. Wohlin 2014) based on a set of four different types of criteria applied on the initial list of papers. This set of criteria consisted of the type of the paper, publication year, publication venue and number of citations. Only scientific papers published in recognized venues with a significant number of citations were included in the final set of papers, which was organized into several categories presented in the following section.

2.2 Primary Studies

As mentioned above, the final list of papers consisted of 81 studies, which were organized in several categories based on their content. Figure 1 shows this list of papers by year of publication. It is worth noting that in each of the years 1996, 1998 and 2005 only one paper was published. In the subsequent years up until and including 2011 the number of publications was slightly increased, while from 2012 to 2018 a significant uptake is observed gradually doubling this figure. This may be attributed to the fact that the general area of Smart Data Processing and Systems of Deep Insight, as a relatively new scientific field or sub-field, did not attract researchers from the very beginning, but gradually the challenges faced gained interest over time.

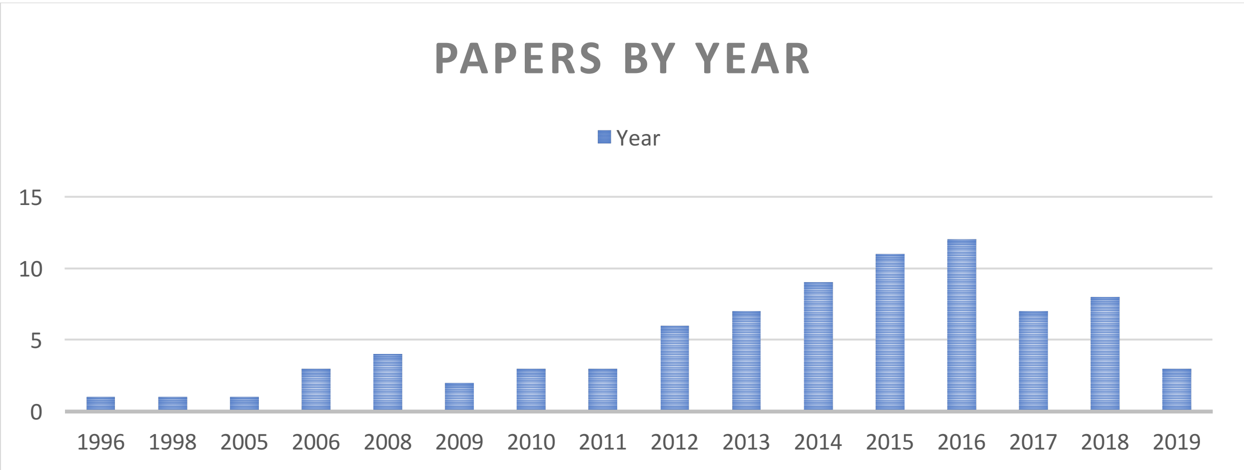

Figure 1. Number of papers examined by year of publication for the smart data processing literature.

Figure 2 shows a categorization of the material gathered in this survey by type: book chapters, conference papers, and journals papers.

In their majority, the articles were published in journals and then in conferences. More specifically, fifty-two (52) of them are journal articles, twenty-two (22) appear in conference proceedings and seven (7) are book chapters.

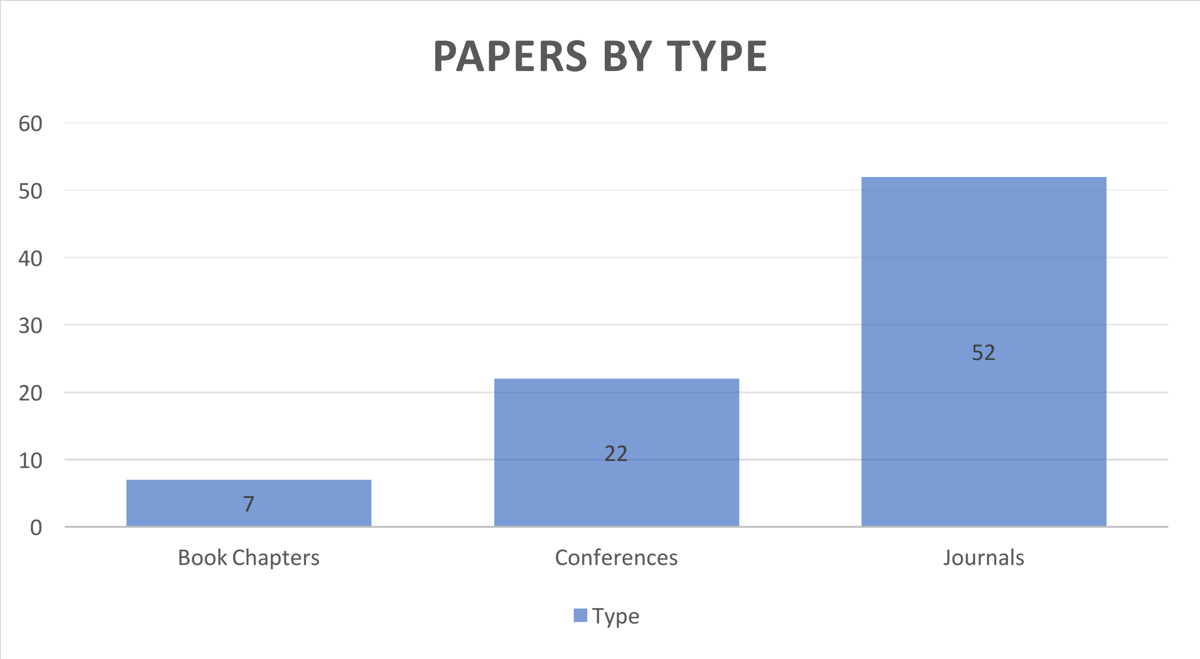

Figure 2. Number of papers examined by publication type.

Based on the content orientation, the papers were split into the three categories formed by the JRAs: (i) Smart Data Processing Systems, (ii) Systems of Deep Insight, and, (iii) Methodology for Data-Centric Services and Applications (see Figure 4). Papers belonging to the first category mainly revolve around data ingestion and data curation, data aggregations, stream-processing, knowledge-based meta data and scalable, reusable and secure processing frameworks (see Figure 3). The second category consists of papers which are mainly involved with large-scale data analytics, cross-correlation and cross functional model, actionable, context-depend deep insights, decision monitoring and next best action and descriptive, predictive cognitive analytics. (see Figure 3). Finally, the last category includes studies which are focused on integrating smart data processing and systems of deep insight in unified data-centric application, on agile methods and DevOps, and on smart data-centric application and data/service evolution (see Figure 3). The following section outlines the main findings of the articles selected in each of the aforementioned categories, focusing on the research challenges faced, their methodology used, their results and the open problems reported.

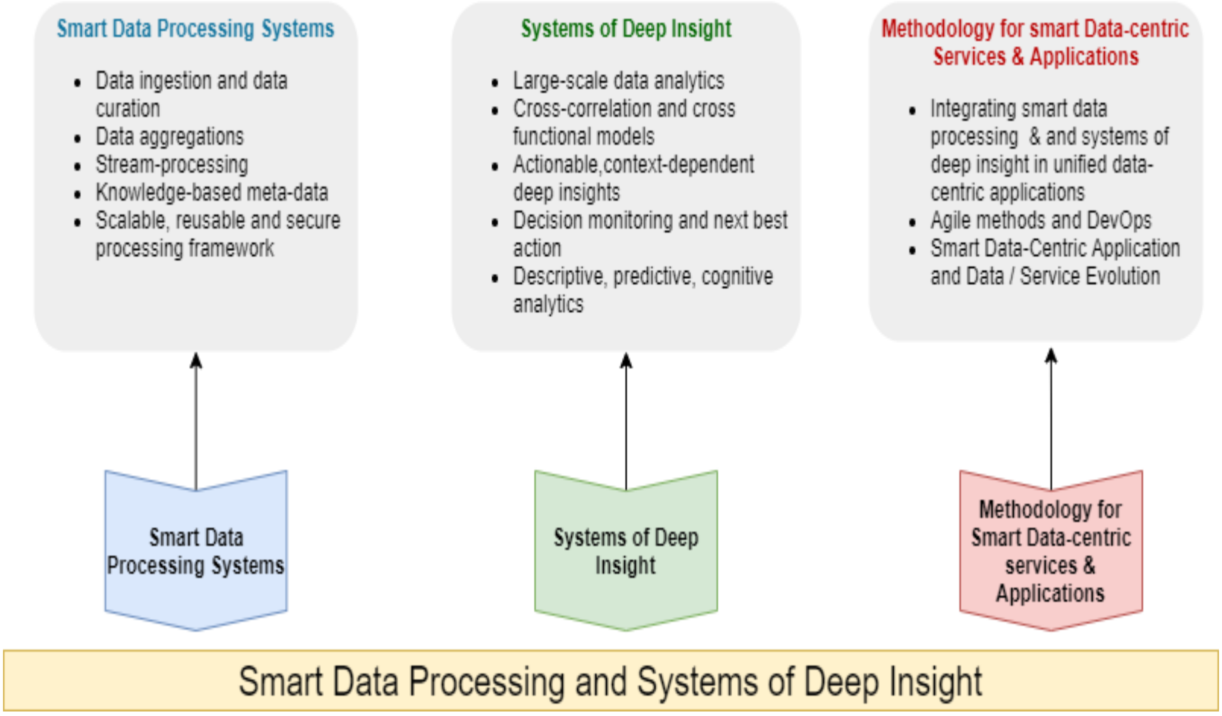

Figure 3. Content and subjects per JRA

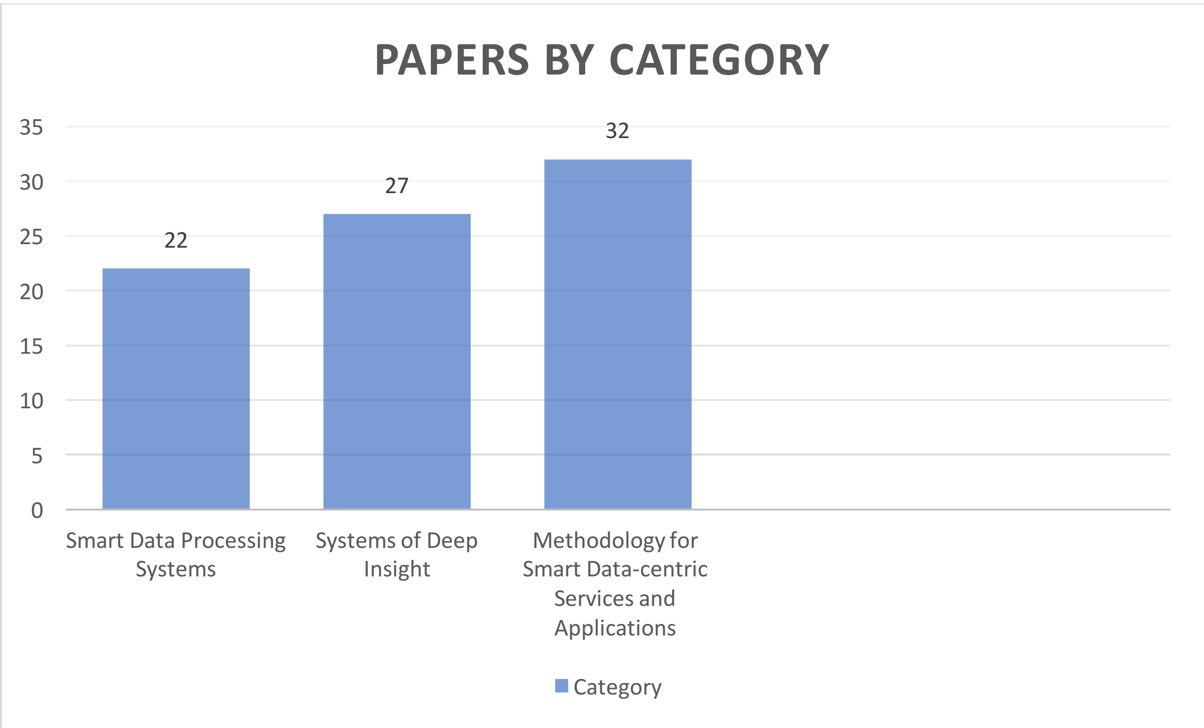

Figure 4. Number of papers examined by JRA

3. Literature Review

3.1 Smart Data Processing Systems

The area of smart data processing comprises the ability to clearly define, interoperate, openly share, access, transform, link, syndicate, and manage data. Under this perspective, it becomes crucial to have various knowledge-based metadata representation techniques to structure data sets, annotate them, link them with associated processes and software services, and deliver or syndicate information to recipients. The Smart Data Processing Systems area can include various topics to fully utilize the aforementioned capabilities such as data ingestion, data aggregation of an enormous variety of structured, unstructured and semi-structured datasets, knowledge-based meta-data representation techniques for the conversion of raw into smart data, data privacy and protection, automated deployment, run-time software performance monitoring and dynamic configuration.

In addition, this area includes adaptive frameworks and tool-suites that support smart data processing by using both data in motion (e.g. data streams from sensors), and data at rest, that rely on advanced techniques for efficient resource management, and partitioning of intensive data workloads across a number of private and public clouds. Smart data processing supports the process and integration of data into a unified view from disparate Big Data sources including Hadoop and NoSQL, data warehouses, sensors and devices in the Internet of Things, social platforms, and databases, whether on-premises or cloud, structured or unstructured and software-as-a-service applications to support Big Data analytics (Yuhanna, 2014).

3.1.1 Smart Data Processing Systems Key Knowledge Areas

The key knowledge areas identified in this category for advancing and enhancing the existing knowledge in Smart Data Processing are as follows:

A. Data Integration Techniques

Most of the work on Big Data integration has been focused on the problem of processing very large sources, extracting information from multiple, possibly conflicting data sources, reconciling the values and providing unified access to data residing in multiple, autonomous data sources. Various studies mainly addressed isolated aspects of data source management relying on schema mapping and semantic integration of different sources Cafarella et al.

(2009), Hassanzadeh et al. (2013) and Venetis et al. (2011). Those studies focused mostly on the construction of a global schema or a knowledge base to describe the domain of the data sources. Web table search is also closely related to data source search. Most of the proposed techniques outlined in Cafarella et al. (2009), Limaye et al. (2010), Das Sarma et al. (2012), Yakout et al. (2012) and Fan et al. (2014) examine user queries and return tables related to specific keywords presented in the query, however, keyword-based techniques fail to capture the semantics of natural language, i.e., the intentions of the users, and thus they can only go as far as giving relevant hits.

B. Source Selection and Knowledge Harvesting

Recent works on source selection Dong et al. (2013) and Rekatsinas et al. (2014) propose various framework that provide the necessary building blocks to derive rigorous time-dependent definitions for data quality metrics, such as coverage and freshness, and statistically models the complex update patterns and data quality changes of different data sources where all sources follow a common schema and focus on a single data domain. An interesting study reported in Rekatsinas et al. (2015) investigates an approach that enables users to discover the most valuable data sources for their applications. This work presents how a system can support the interactive exploration of different sets of data sources, allowing the user to truly understand the quality and cost trade-off between different integration options.

Other prominent example of such studies can be found in recent efforts extracting entities, relationships and ontologies from the Web to build general purpose knowledge bases, such as Bollacker et al. (2008), the Google knowledge graph Dong et al. (2013), ProBase Wu et al. (2012), and WebChild Tandon et al. (2014) in order improve web-applications.

Finally, the work in Weikum et al. (2016) presents knowledge-harvesting techniques that developing data extracting methods for noun phrase parsing to construct large knowledge bases from various Internet data sources.

3.1.2 Smart data processing systems literature review

Common fields of data processing systems are semantic models, structured data configurations and ontologies. Various papers incorporate ontologies to tackle data processing issues:

Lanzenberger, Sampson, and Rester (2010) examined an enormous number of ontology visualization tools to identify solutions for dealing with the complexity of large ontologies. Their work was a starting point to demonstrate the usefulness of Information Visualization techniques, aimed to boost the adoption of ontologies in common Web applications.

Yang, Dong, and Miao (2008) adopted an expressive OWL (Web Ontology Language) ontology language and a SWRL (Semantic Web Rule Language) rule language to model product configuration knowledge which have the advantage of reusing configuration models that is crucial, considering incremental changes and updates on products due to new technology advances.

Roda and Musulin (2014) propose an ontology-based framework for IDA (Intelligent Data Analysis) which is based on a knowledge model composed by existing ontologies, the Semantic Sensor Network ontology (SSN) and the SWRL Temporal Ontology (SWRLTO), and a new developed one, the Temporal Abstractions Ontology (TAO). They demonstrate their framework by using it in a chemical plant case study to show how complex temporal patterns that combine several variables and representation schemes can be used to infer process states and/or conditions.

The work of Petersen et al. (2017) mentions that the digitization of the industry requires information models that describe assets of companies to enable the semantic integration and interoperable exchange of data. Their proposed model is centred around machine data and describes all relevant assets, key terms and relations in a structured way. They evaluated their approach with stakeholders on two case studies. While the stakeholders find the advantages of semantic technologies appealing, the lack of ready-to-use business solutions, industrial ontologies and available IT personnel is halting their efforts to move forward.

Drabent et al. (2009) firstly outline the current state of the Semantic-Web stack and its components, and then discuss the open issues in combining rules and ontologies before defining a combined rule and ontology knowledge-base two-step redact in which, as a first step, the ontology predicates are eliminated under the open-world assumption (OWA) and, as a second one, the negated logic-programming predicates under the closed-world assumption (CWA).

Mehdi et al. (2017) reports that industrial rule-based diagnostic systems are often data-dependant in the sense that they rely on specific characteristics of individual pieces of equipment. This dependence poses significant challenges in rule authoring, reuse, and maintenance by engineers. That work addresses the aforementioned problems by proposing a semantic rule language, sigRL, where sensor signals are first class citizens. Their evaluation shows that up to 66% of the time is saved when employing ontologies and that execution of semantic rules is efficient and scales well to real-world complex diagnostic tasks.

Cuenca, Jim, and Mehdi (2016) describe the outcomes of an ongoing collaboration between Siemens and the University of Oxford, with the goal of facilitating the design of ontologies and their deployment in applications. They present SOM, a tool that supports engineers in the creation of ontology-based models and in populating them with data.

Bock et al. (2010) show how to combine ontological and model-based techniques in languages that facilitate collaborative design exploration. The proposed approach uses ontology to capture alternative designs and incremental refinements that meet requirements and earlier design commitments. In this work model-based techniques are applied to develop more powerful, engineering-friendly languages for using ontology.

Jorgesen (2008) introduced fundamental concepts of product configuration. The introduction and implementation of product configuration demand a systematic way of thinking in constructing, documenting, and maintaining the configurable products. This can be achieved by defining a product family model as a model of a set of possible products.

Li, Xie, and Xu (2011) state that product knowledge has played an increasingly significant role in new product development process especially in the development of One-of-a-Kind products. Their paper provides a comprehensive review on the recent development of knowledge-based systems (KBS), methods and tools in supporting rapid product development.

The Internet of Things (IoT) is nowadays a vital source of data, both in terms of volume and frequency of production. Quite a few studies are devoted to the study of problems pertaining to the collection, structuring, processing and presentation of IoT data towards the development of new applications and services:

Lee and Lee (2015) firstly identify the mostly widely used IoT technologies that are essential in the deployment of successful IoT-based products and services and then discuss the three IoT categories for enterprise applications to enhance customer value.

Qin et al. (2016) review the main techniques and state-of-the-art research efforts in IoT from data-centric perspectives, including data stream processing, data storage models, and complex event processing. This paper covers investigations on data models, search and event processing, and present the potential of IoT applications in smart cities, environment monitoring, health and energy home.

Data structuring, organisation and fast processing has also gained significant interest during the last decades, with studies investigating a rich number of relevant issues:

Over the years a rich ecosystem emerged around Hadoop comprising tools for parallel, in-memory and stream processing. Luckow et al. (2015) survey use cases and applications for deploying Hadoop in the automotive industry and argue about the need to develop automotive applications and requirements for data discovery, integration, exploration and analytics.

Dean and Ghemawat (2008) outline the novel programming model MapReduce, which has been successfully used by Google for many different purposes. The authors attribute this success to several reasons: Firstly, the model is easy to use, even for programmers without experience with parallel and distributed systems. Secondly, a large variety of problems are easily expressible as MapReduce computations. Thirdly, an implementation of MapReduce has been developed that scales large clusters of machines.

Guerrero et al. (2017) propose a heterogeneous data source integration based on IEC (Electrotechnical Committee) standards and metadata mining. The system includes several data mining tools to model information for classification, outlier detection, pattern detection, forecasting, or information retrieval based on the level of importance established by metadata mining process.

Erkin et al. (2013) present recent and ongoing research in the field of privacy protection for smart grids, where individual smart meter measurements are kept secret from outsiders, including the utility provider itself, while processing private measurements under encryption is still feasible. The authors focus particularly on data aggregation, which demonstrates the major research challenges in privacy protection for smart grids and conclude that researchers should invest more in cryptography.

Miloslavskaya and Tolstoy (2016) firstly state that a data lake holds a vast amount of raw data in its native format and then define fast data as a time-sensitive structured and unstructured “in-flight” data that should be gathered and acted upon right away. The authors conclude that not all big data is fast, as well as not all fast data is big.

Khine and Wang (2018) argue that a data lake is one of the arguable concepts appeared in the era of big data. The idea of a data lake is originated from business field instead of the academic. As data lake is a newly conceived idea with revolutionized concepts, it brings many challenges for its adoption. However, the potential to change the data landscape makes the research on data lakes worthwhile.

Fang (2015) discusses the concept of data lakes and shares the author’s thoughts and practices on the subject. The main goal of the paper is to examine and provide answers to a series of questions: What is a data lake? How does it help with the challenges posed by big data? The author concludes that the data warehouse is a wise choice for a company dealing with big data challenge and outline the best practices of data lake implementations.

3.2 Systems of deep insight

The area of Systems of Deep Insight focuses on analytic solutions that enable optimization of asset performance in smart data processing systems and is geared by systems that turn data into insights. This category relies on smart data processing and integration techniques that utilize data in engagement and records systems. In addition, an analysis on how to shift through the data to discover new relationships and patterns is briefly examined.

3.2.1 Systems of Deep Insight Key Knowledge Areas

The key knowledge areas outlined below are essential for advancing and enhancing existing knowledge in Systems of deep insight:

A. Predictability and Prescriptiveness

Predictability and Prescriptiveness are the two main characteristics that systems of deep insight aspire. The key idea behind predictability is that learning can be thought of as inferring plausible models to explain observed data. Probability theory provides a solid framework where a decision may depend on the amount of uncertainty. The dominant paradigm for representing such probabilistic models with variants is examined in Koller, D. & Friedman. (2009). Traditionally, the problem of integration over the various plausible outcomes has been considered a source of high computational burden. However, recent advances in the field, including black-box variational approximations Rui et al. (2016) and stochastic gradient Markov chain Monte-Carlo (SG-MCMC) Chen et al. (2016) have completely ameliorated these issues, by rendering Bayesian probabilistic models amenable to large-scale data analytics applications. In addition, probabilistic programming constitutes a recent culmination on the aforementioned research efforts, allowing the use of computer programs to represent Bayesian probabilistic models. There is a growing number of probabilistic programming languages currently under active development; Stange (2019), Infer.NET Minka et al. (2016), and Edward Tran et al. (2016) are only few such examples.

Prescriptiveness refers to the ability to prescribe an action so that the decision-maker can take insight information and act. Prescriptive analytics requires a predictive model able to predict the possible consequences based on different choices of actions. In the context of systems of deep insight, Li et al. (2010) consider the problem of planning an action that maximizes the potential of short-term gain, based on historical data, while gathering new information for improving goodness between actions and long-term effects. This dilemma is typically formulated as a contextual multi-armed bandit problem where each arm corresponds to one possible course of action. The optimal strategy is to use the arm with the maximum expected reward as regards contextual information on each trial, and then to maximize the total accumulated reward for the whole series of trials.

Recently, a series of algorithms for contextual multi-armed bandit problems have been reported with promising performance under different settings, including unguided exploration (e.g., ε-greedy Tokic (2010) and epoch-greedy Langford and Zhang (2007) and guided exploration (e.g., Thompson Sampling Chapelle and Li (2011)). These existing algorithms take the contextual information as the input and predict the expected reward for each arm, assuming the reward is invariant under the same context. Finally, other methods that capture the time varying behaviours of the reward in contextual multi-armed bandit problems have also been proposed in Chunqiu (2016).

B. Monitoring, Reconfiguration & Self-adaptation

Monitoring of data support infrastructure, including cloud computing and service delivery environments, requires standardized metrics that enable efficient controlling of its resources towards improved quality. Examples of metrics based on the current literature review include the following: delivery cycle time, mean time to detect problems and weaknesses, mean time to repair them, quality at the source, etc. (Aceto et al. 2013).

Resource management adjusts the availability of resources. Reconfiguration or self-adaptation may take the form of a proactive process or of a set of reactive tasks. In the former case the upcoming traffic peaks and workloads are estimated or predicted upfront and actions are taken based on specific metrics before the actual need commences. In the latter case reaction is realized in response to observing or measuring traffic peaks and workload demands at the time of occurrence.

Particular interest to this JRA are current approaches outlined in Ardagna et al. (2012), Addis et al. (2013) and Ardagna et al. (2014) that propose proactive actions to ensure Qulaity of Service (QoS) delivery according to SLA (Service Level Agreement) stipulations. The cornerstone of the relevant research activities is a SLA profile. Based on this profile methods and algorithms are built to monitor the level of QoS offered, juxtaposed with the provisions of the SLA. The outcome of this monitoring is the collection of information that enables the management of resources to handle difficult cases (e.g. in terms of performance, security, precision etc.) where the risk of SLA violation may be avoided by predicting its likelihood and assessing its impact.

3.2.2 Systems of Deep Insight Literature Review

Various studies belong to this category, investigating topics which include big data, services, to Cyber Physical Systems (CPS), business intelligence, machine learning techniques and algorithms, and various applications to real-world problems which involve models and systems that provide insights for decision support, optimization and control.

Barnaghi, Sheth, and Henson (2013) describe the Big Data issues in the Web of Things (WoT), discuss the challenges of extracting actionable knowledge and insights from raw sensor data, and introduce the theme articles in this special issue. The authors demonstrate different steps that can be envisaged for efficient processing and for making use of WoT data.

Delen and Demirkan (2013) provide a conceptual framework for service oriented managerial decision-making process, and briefly explain the potential impact of service-oriented architecture (SOA) and cloud computing on data, information and analytics. The authors believe that their proposed approach to service-oriented data, information and analytics in the cloud will create great opportunities, as well as many challenges.

Stojmenovic (2014) explores Cyber Physical Systems (CPS) beyond the M2M (Machine to Machine) concept before describing a number of particular use cases that motivate the development of the M2M communication primitives tailored to large-scale CPS. The author argues that there is a need to design M2M communication primitives able to scale to thousands and trillions of M2M devices, without sacrificing solution quality.

Larson and Chang (2016) examine the application of Agile methodologies and principles to data-driven business intelligence delivery and discuss how these methodologies also changed with the evolution of business intelligence. In addition, the authors address how Agile principles and practices have evolved with business intelligence as well as their challenges and future directions.

Wang, et al. (2019) present a new deep learning-based machine vision inspection method to identify and classify defective product without the loss of accuracy. More specifically, firstly a Gaussian filter is utilized on an acquired image to minimize the random noise and secondly, a region of interest (ROI) is conducted based on the Hough transformation to remove the unrelated background, thereby offloading the computational burden of the subsequent identification process. The experimental study on defective bottles inspection demonstrates the usefulness of the proposed method.

Lee et al. (2014) discuss the trends of manufacturing service transformation in big data environments, as well as the readiness of smart predictive informatics tools to manage big data, thereby achieving transparency and productivity. The objective of the paper is to review how current manufacturing industries evolve for the upcoming industrial big data environment, and to propose the key technology for sustainable innovative service.

The work of Yan et al. (2011) sets the ground for research on home power management systems optimization as regards to the privacy of customer power usage behaviors**.** The performance of the reading data aggregation and dispatch has been analyzed subject to the HAN setting. The levels of security were discussed qualitatively, focusing on the secrecy of pseudo-random spreading codes and circuit shift. Simulation results demonstrated the advantage of the proposed scheme over the traditional BSS approach.

Kim (2017) presents a new transactional scheduler, called partial rollback-based transactional scheduler (or PTS), for a multi-versioned DTM (Distributed Transactional Memory) model. The model supports multiple object versions to exploit concurrency of read-only transactions, and detects conflicts of write transactions at an object level. PTS's design shows that partial rollback-based scheduling is a viable strategy for transactional processing in in- memory data grids.

Mahdavinejad et al. (2018) assess various machine learning methods that deal with the IoT data challenges extracted from a smart city use case. The key contribution of this study is the presentation of a taxonomy of machine learning algorithms explaining how different techniques are applied to the data in order to extract higher level information.

Chen et al. (2012) initially argue that business intelligence and analytics (BI&A) has emerged as an important area of study for both practitioners and researchers, reflecting the magnitude and impact of data-related problems to be solved in contemporary business organizations. They continue their work by reporting a bibliometric study of critical BI&A publications, researchers, and research topics based on more than a decade of related academic and industry publications.

Farid et al. (2016) present CLAMS, a system to discover and enforce expressive integrity constraints from large amounts of lake data with very limited schema information (e.g., represented as RDF triples). CLAMS has been deployed in a real large-scale enterprise data lake compromising 1.2 billion triples and was able to spot multiple obscure data inconsistencies and errors early in the data processing stack, providing huge value to the enterprise. This paper shows how CLAMS holistically combines the signals from diverse constraints spanning over multiple datasets and utilize user feedback to obtain accurate repairs.

Aceto et al. (2013) provide a survey on Cloud monitoring. The authors start by analysing motivations for Cloud monitoring, providing also definitions and background for the following contributions. Then, they analyse and discuss the properties of a monitoring system for the Cloud, the issues arising from these properties and how such issues have been tackled in literature.

Charest and Delisle (2006) propose the realization of a hybrid intelligent data mining assistant, based on the synergistic combination of both declarative (Description Logic) and procedural (SWRL Rules) ontology knowledge in order to empower the non-specialist data miner throughout the key phases of the CRISP-DM data mining process. The authors successfully present some novelty features their intelligent DM assistant attempts to provide by combining both declarative and procedural ontology knowledge. Furthermore, the use of the DM ontology provides a natural extension to the existing CBR (Case-Based Reasoning) for addressing the need for “deeper” knowledge to empower the data miner.

Lee et al. (2013) discusses the principles of predictive manufacturing system as a strategy to allow the manufacturing industry to increase competitiveness through a highly transparent and worry-free manufacturing process, as well as an analytic framework that can be implemented using a coupled model approach to unravel and measure uncertainties in certain industries.

Yu and Boyd (2016) outline a general-purpose flexible in-memory indexing technique based on multi-level key ranges, which can be easily adopted into existing systems with B+-tree, ISAM or data list of sortable keys to make the indexing smarter.

Saldivar et al. (2016) present a k-means cluster approach used to manage relevant big data. The identification of patterns from big data is achieved with a cluster method and with the selection of optimal attributes using genetic algorithms. The final outcomes of this work present that big data analytics (nodes) help to visualize the influence of product characteristics and to cluster customer needs and wants.

Wang et al. (2018) present a comprehensive survey of commonly used deep learning algorithms and discuss their applications toward making manufacturing “smart”. Specifically, a deep learning enabled advanced analytics framework is proposed to meet the opportunistic need of smart manufacturing. Deep learning provides advanced analytics and offers great potentials to smart manufacturing in the age of big data. By unlocking the unprecedented amount of data into actionable and insightful information, deep learning gives decision-makers new visibility into their operations, as well as real-time performance measures and costs.

Chungoora et al. (2013) propose a hybrid approach that combines federated and multi database techniques, which provide the most feasible avenue for large scale integration. Under the proposed architecture, the individual data site administrators provide an augmented export schema specifying knowledge about the sources of data, their structure, their content and their relationships. This knowledge is used to generate a partially integrated, global view of the data.

Azvine et al. (2006) firstly discuss the issues and problems of current business intelligence systems, and then outline their vision of real-time business intelligence. In addition, they present a list of emerging technologies that are being developed within the research program of British Telecommunications (BT) plc, which could contribute to the realisation of real-time business intelligence.

Ben et al. (2005) discusses issues and problems of current business intelligence systems, and then outlines our vision of future real-time business intelligence. Moreover, present a list of emerging technologies which could contribute to the realisation of real-time business intelligence and some examples of applying them to improve BT’s systems and services. Ben Azvine, Cui, and Nauck (2005) presents the future RTBI infrastructure will include the following elements: 1) static data warehouses and dynamically user- configurable data shopping malls, 2) meta-data information for the whole enterprise, 3) taxonomies and ontologies for describing contents and providing semantic content information, 4) information about the context of data sources, 5) advanced ETL tools for gathering and feeding data to analytical, 6) feedback mechanisms to operational systems.

Polyvyanyy et al. (2017) proposes the Process Querying Framework, which aims to guide development of process querying methods. given a process repository and a process query that specifies a formal instruction to manage the given repository, the corresponding process querying problem consists of implementing the instruction on the repository.

Sahay and Ranjan (2008) examine the need for real-time business intelligence (BI) in supply chain analytics. The authors focus their interest on the necessity to revisit the traditional BI concept that integrates and consolidates information in an organization to support firms that are service oriented and seeking customer loyalty and retention.

Tang et al. (2018) describe the vision of smart shop-floor based on the notion of Industry 4.0 that denotes technologies and concepts related to Cyber-Physical Production Systems (CPPS). The experimental results prove that intelligent manufacturing paradigms aligning with smart shop-floor enable agile reaction to disturbances and maintenance of high production performance.

Denno et al. (2018) present a methodology, called production system identification, to produce a model for a manufacturing system from system's operation logs. The model produced is intended to aid in making production scheduling decisions. The proposed methodology is evaluated on an automotive assembly system concluding that it is possible to use log content to produce a model useful to production control tasks, such as line balancing and job sequencing.

Zhong et al. (2015) propose a holistic Big Data approach to excavate frequent trajectory from massive RFID-enabled shop-floor logistics data with several innovations highlighted in order to deal with existing methods which are not suitable for removing noises due to the highly complex and specific characteristics of RFID Big Data.

Borkar et al. (2012) propose the use of recursive queries to program a variety of machine learning algorithms instead of creating a new system for each specific flavor of machine learning task, or hard-coding new optimizations. By utilizing this approach, database query optimization techniques can be used to identify effective execution plans, which can be executed on a single unified data-parallel query processing engine. The authors demonstrated that their approach can offer a plan tailored to a given target task and data for a specific machine resource allocation.

Rusitschka et al. (2010) present a cloud computing model for managing the real-time streams of smart grid data using real-time information needs The Smart Grid Data Cloud is suitable for liberalized energy markets with a data clearing house concept, large vertically integrated utilities, as well as associations of transmission system operators, such as in the ENTSO-E.

Lu and Wen (2014) propose a minimum-cost-forwarding-based asynchronous distributed algorithm to find the optimal placement for the data aggregation service tree with optimal cost of in-network processing. The authors demonstrate that the proposed algorithm has less message overheads than the synchronous algorithm (Sync).

Smart grid data analytics play a critical role in the business and physical operations of delivering electricity and managing consumption. Even though utilities start from a difficult position as there is need to integrate data analytics into the enterprise, data science is critical to modernize the grid. Stimmel (2016) demonstrate the critical role that smart grid data analytics bring to the electricity business.

Papazoglou et al. (2015) present a production architecture which lowers the barrier for entrepreneurs to design novel products and processes and develop manufacturing software that could be plugged into the SMN platform for easy access by multiple users to enable collaborative manufacturing of new products and response to product demand.

Roh et al. (2019) perform a comprehensive study of data collection from a data management point of view and discuss interesting data collection challenges that remain to be addressed by the research community.

O'Leary (2014) examines the notion of the Big Data Lake and contrast it with existing solutions (data warehouses) to discuss the risks of the Emerging Lake concept, and investigate the embedding of different artificial intelligence and crowdsourcing (human intelligence) applications into that lake. The final results present that the emerging conceptual vision of the lake is able to integrate and analyze multiple data sources in a single table captured as part of in-memory computing.

3.3 Methodology for Smart Data-Centric Services and Applications

The Methodology for Smart Data-Centric Services and Applications category examines smart application development techniques by providing a methodology that interlocks elements of smart data processing and systems of deep insight to alleviate complexity and the effect of changes, to speed up the entire software development/deployment process for smart applications.

Smart services and applications lean on support provided by smart data processing systems and systems of deep insight (the two previous categories), respectively, where the data must be gathered on an ongoing basis, analysed, and then provide direction to the business regards to any appropriate actions to take, thus providing value. Data in smart applications may originate from a variety of sources, including intelligent sensors and devices that are transmitting data (popularly called the Internet of Things) and by other sources of semi-structured and structured data.

Smart applications involve data-intensive software development in conjunction with effective analytics techniques that utilize/produce data. Data analysis challenges are related to processing and generating insights from the massive amount of data stored. In addition, the software development process in smart applications started to be considered as a socio-technical arrangement, where organizational and human aspects play a key role in the business.

3.3.1 Methodology for Smart Data-Centric Services and Applications Key Knowledge Areas

The key knowledge areas outlined below are essential for advancing and enhancing existing knowledge in this category.

A. Distributed Application Development Methodologies

Distributed application development encompasses modern design principles and proven practices to facilitate the development tasks and provide developers with a roadmap for building robust and correct distributed services and applications. An important element of distributed application development is to achieve a clean separation between design time and runtime actions.

The Agile Manifesto has spawned many approaches and methods relevant to the development of distributed applications. Among the others, Scrum is certainly very popular and successful Deemer et al. (2012). Similarly, Application Lifecycle Management (ALM) suite is a class of products that offers mechanisms for automating some tasks and for connecting managerial tasks with software development activities Chappel (2008). Agile development is closely related to DevOps, which is the evolving integration between the software developers who build and test applications, and the IT teams that are responsible for deploying and maintaining IT systems and operations Debois (2001). DevOps can help any organization dramatically speed up application and delivery cycles. Recently the authors in Tamburri et al. (2014) have recognised the importance of organizational social structures to identify the ones that are suitable to the software engineering domain for developing distributed application.

Finally, it is worth mentioning that the concept of Master Data Management (MDM) is closely related to the DevOps approach. MDM is a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise’s official shared master data assets. Master data describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies, etc. (Radcliffe, 2009). Such considerations are important when developing a methodology for smart data-centric applications and services.

B. Domain-specific Languages for Data-intensive Applications

Traditionally, computing models such as MapReduce Dean and Ghemawat (2008), were proposed to support data-intensive applications. Although such models suited massive-scale data processing, they permit limited application logic complexity Kalavri and Vlassov (2013) Domain Specific Languages (DSLs) can be employed to circumvent such problems. DSLs offer pre-defined abstractions to represent concepts from the application domain. DSL compilers may optimize the code written for the specific domain. DSLs ease the implementation of analytics and machine learning algorithms with the use of high-level abstractions or reusable pieces of code that hide low-level details from software engineers letting them focus on the main problem at hand.

Languages like OptiML (Chafi et al., 2011) enable machine-learning algorithms to take advantage of parallelism by bridging the gap between machine learning and heterogeneous Big Data hardware infrastructure. OptiML is a declarative, statically-typed textual programming language, in which variables have their types specified before execution. OptiML operations support parallel executions (using the MapReduce programming model) in heterogeneous machines but it lacks support for a distributed environment or executions in the cloud.

In addition, ScalOps (Wu et al., 2015) is another example of declarative, statically-typed textual programming language DSL, with the goal of enabling machine learning algorithms to run on a cloud computing environment and overcome the lack of iteration limitation of the traditional MapReduce programming model. To support iterations in MapReduce, ScalOps introduces an enhanced version of the programming model called Map-Reduce-Update where a map function receives read-only global state values and is applied to training data points in parallel, while a reduce function aggregates the output of the map function (Borkar, 2012).

3.3.2 Methodology for Smart Data Centric Services and Applications Literature Review

This section is devoted to papers that propose or utilize methods, techniques or steps that aim to create and execute smart data centric services or applications, as well as their effectiveness in real-world case-studies.

Chungoora et al. (2013) argue that the novel combination of the MDA approach with formal ontology driven specifications based on core and specialised domain ontologies can provide a route to improve knowledge sharing across product design and manufacturing. The methodology applied in this work was evaluated on an industrial use case based on the design and manufacturing of aerospace parts. The results from the experimental investigation of the

IMKS model-driven concept were promising in facilitating interoperation across systems for improved knowledge sharing.

Pang et al. (2015) present an innovative Data-Source Interoperability Service (DSIS) that serves as a middleware for providing a querying and information integration service for heterogeneous data sources. The DSIS applies software agent technology that is capable of accomplishing tasks in an autonomous way without human intervention.

Lu and Xu (2019) present a generic cloud-based manufacturing equipment architecture based on cyber-physical systems and big data analytics to solve a particular problem that cloud-based equipment faces: The proposed system can cache unprocessed machine logs onto the local machine while waiting for machine data to be submitted to the cloud. An industry implementation in a world-leading cloud provider confirmed that the proposed architecture can successfully enable on-demand manufacturing services provisioned via the Internet, which can be extended to businesses that endeavour to transform legacy production systems into cloud-based cyber-physical production systems.

Petersen et al. (2016) propose an enhanced semantic model which enables views spanning from the high level of supply chains to the low level of machines on a shop-floor. The model includes a mapping to relational production databases to support federated queries on different legacy systems in use. This work is focused on a production line use-case, demonstrating that it can be used for typical factory tasks, such as assembly line identification or machine availability checks.

O'Donovan et al. (2015) present an industrial big data pipeline architecture, focused on equipment maintenance in large-scale manufacturing that differs from traditional data pipelines and workflows as it has the ability to seamlessly ingest data from industrial sources, co-ordinate data ingestion across networks using remote agents, and automate the mapping and cleaning process for industrial sources of time-series data. The contributions and findings of this research work are important for facilitating big data analytics research in large-scale industrial environments.

Zhang et al. (2018) propose a CPN-MIASS for smart factory that aims to establish a systematic graphic-based modelling approach for capturing manufacturing information that can assist the general operators in monitoring and controlling the real-time manufacturing process easily and dynamically.

Cafarella et al. (2009) present Octopus, a system that combines search, extraction, data cleaning and integration, and enables users to create new data sets from those found on the Web. Octopus executes some operators automatically, but always allows the user to provide feedback and correct errors.

Garetti et al. (2013) introduce an innovative open knowledge-based manufacturing control system solution based on ontology knowledge and Service Oriented Architecture (SOA) approaches, which allow the control to be automatically customized by the ontology information on the physical system.

Tao et al. (2014) present that Internet of Things (IoT) and Cloud Computing (CC) have been widely studied and applied in many fields, as they can provide a new method for intelligent perception and connection from M2M (including man-to-man, man-to-machine, and machine-to- machine), and on-demand use and efficient sharing of resources, respectively. The authors explained how to realize the full sharing, free circulation, on-demand use, and optimal allocation of various manufacturing resources and capabilities, as well as the applications of the technologies of IoT and CC in manufacturing. The main contributions of this paper are multiple: An overview of the applications of IoT and CC in the manufacturing field; the potential of advanced technologies, such as IoT and CC, for addressing the bottlenecks faced by the existing AMSs is investigated; the introduction of a CC- and IoT-based CMfg system; the relationship among CMfg, IoT, and CC; and, finally, the technologies systems for realizing the CCIoT-CMfg.

Wang et al. (2018) examine the development, architectural design and component functionality of big data analytics by analysing 26 big data use cases in healthcare in order to identify the data analytics capabilities in the health domain. Their final outcomes identified the following capabilities: patterns, unstructured data, decision support, predictive, and traceability.

Terkaj and Urgo (2015) identified the need for a common framework to support the interoperability and exploitation of different actors with different competences and expertise within a factory, and addressed the use of an ontology-based model in a production system to support the construction of a performance evaluation model. This study focused on providing a complete data model for production systems by linking the static characterization of production resources with evaluation activities.

Jardim-Goncalves et al. (2014) present NEGOSEIO, a framework which enables service-based interoperability between parties, closely integrated with semantics and business understanding via the use of reference ontologies in the quest for achieving a stronger interoperability liaison. The authors proposed a framework that offers negotiation mechanisms to support the sustainability of interoperability in business-to-business interactions, in networked enterprise environments. The use of NEGOSEIO in case studies reduced the decision time and provided suggestions for better solutions.

Giese et al. (2015) integrate a user-oriented query interface, with semi-automated managing methods, new query rewriting techniques, and temporal and streaming data processing approaches in one platform to deal with the current problems in ontology-based data access systems pertaining to installation overhead, usability, scalability, and scope. Optique offers a single point of entry for administrative tasks, managing mappings and ontologies, as well as visual components, that let users interact with big data to satisfy their information needs.

Lin et al. (2012) present a Global Decision Support System for Small or Medium Enterprise (SM/E) which enables different functional units to analyse decision making and propose goals that are in favour of their respective functional units' performance. This enables the managers of local manufacturers to efficiently conduct collaborative decision-making activities in relation to other participants of collaborative manufacturing.

Giovannini et al. (2012) state that the required cultural shift needs actions that will involve deeply software and hardware aspect of the manufacturing processes. This paper addresses more the software part of this challenge by proposing a product centric ontology, in which concepts of product, processes and resources are associated with functions and sustainable manufacturing knowledge. The results showed that the proposed model can exploit captured knowledge to propose design and manufacturing process changes.

Wang et al. (2016) present a smart factory framework that incorporates industrial network, cloud, and supervisory control terminals with smart shop-floor to cooperate with each other. Simulation results assess the effectiveness of the proposed negotiation mechanism and presents deadlock prevention strategies.

Lemaignan et al. (2006) present a manufacturing upper ontology, aimed to draft a common semantic net in manufacturing domain namely MASON (Manufacturing's Semantics Ontology). The authors explain how ontologies play a central role in intelligent manufacturing and enable fluent and consistent flows of data to offer mature tools to deal with challenges like the ease integration with other implemented information systems.

Bashir and Gill (2017) present an IoT Big Data Analytics (IBDA) framework used for storage and analysis of real-time data generated from IoT sensors deployed inside a smart building to fill the research gap in the Big Data Analytics domain. The initial results indicate that the proposed framework is a good fit for the purpose it was designed for and seems useful for IoT-enabled Big Data Analytics for smart buildings.

CPS is a system of collaborating computational entities which are in intensive connection with the surrounding physical world and its on-going processes, providing and using, at the same time, data-accessing and data-processing services available on the internet. Currently, a precursor generation of CPS can be found in areas as diverse as aerospace, automotive, civil infrastructure, chemical processes, energy, healthcare, transportation and manufacturing. Lee et al. (2015) describe a CPS architecture in a shop-floor for intelligent manufacturing and identify three key enabling technologies for CPS implementation: interconnection and interoperability among different devices; multi-source and heterogeneous data acquisition, integration, processing and visualization; and intelligent decision-making based on knowledge acquisition and learning methodology.

Harjunkoski and Bauer (2014) assess the feasibility of using the ISA-95 standard for transferring input/output data to and from a scheduling solution. They also propose on how to apply the standard for a relatively generic set of scheduling problems.

Sakr and Elgammal (2016) propose an integrated and comprehensive framework for big data analytics services in smart healthcare networks, namely SmartHealth, which acts as a roadmap for the research in the area of big data analytics in smart healthcare applications. The authors suggest that the increasing volumes of information gathered via patient monitoring systems is a continuous phenomenon that leads to large datasets ready to be used for big data analytics services and utilize its various applications in the healthcare domain.

Sun et al. (2015) present the implementation of an indoor localization system that aims to implement tracing and tracking of workers and working parts in a future smart factory.

Gharaibeh et al. (2017) provide a data-centric perspective, that describes the fundamental data management techniques employed to ensure consistency, interoperability, granularity and reusability of the data generated by IoT in smart cities. The authors conclude that the lack of a standard definition of a smart city eventually results in various shortcomings in many facets of a smart city.

Lee et al. (2015) focus on existing trends in the development of industrial big data analytics and CPS and introduce a systematic architecture for applying CPS in manufacturing called 5C. The 5C architecture includes all necessary steps to fully integrate cyber-physical systems in the manufacturing industry. In addition, a case study for designing smart machines through the 5C CPS architecture is presented that shows the integration of the 5C architecture for processing and managing a fleet of CNC sawing machines which are commonly used in manufacturing.

Dong et al. (2012) propose a randomized solution for selecting sources for fusion that can efficiently estimate fusion accuracy and select the set of sources that maximizes the profit. The authors present its effectiveness and scalability on both real-world data and synthetic data.

Papazoglou and Elgammal (2018) propose a new manufacturing paradigm called manufacturing blueprints, which allows manufacturers to move from a traditional product-centric business model to a fully digital, knowledge-based and service-centric one. Using this paradigm, manufacturers are able to combine manufacturing and equipment data and knowledge, production systems and processes to form a smart manufacturing network to diversify products and build new markets.

Zhang et al. (2012) describe a new Machine-to-Machine (M2M) communication paradigm, namely Cognitive M2M (CM2M) communication, that uses the cognitive radio technology in M2M communications. A CM2M communications architecture for the smart grid is presented in this work, which also proposes an energy-efficiency driven spectrum discovery scheme.

4. Challenges and Open Research Problems

Several challenges and open research problems were addressed in the studies reviewed, which span more than one of the scientific areas presented in section 3. Therefore, these are organised in classes that describe the nature of the challenge or problem rather than the area they belong to. In general, the following classes were identified:

I. Optimization, Decision Support, Prediction
II. Utilization of artificial or computational intelligence
III. Enhancement of experimentation
IV. IoT infrastructure, technology and applications
V. Data modeling and characteristics
VI. Cloud environment
VII. Smart data processing
VIII. Standardization

A brief description follows of the most significant challenges reported in each of the classes mentioned above.

4.1 Optimization, Decision Support, Prediction

Several challenges have been reported in this category. The most significant ones are the following:

- There is need for new algorithms and techniques to find feasible and optimal solutions of product configuration in terms of some objective (Yang et al. 2008).
- Automatic data mining techniques must be in place based on the results of metadata mining in order to increase the accuracy of generated models. (Guerrero et al. 2017)
- Machine health prediction is a future target that will reduce machine downtime, and the prognostics information will support the optimization of manufacturing management, maintenance scheduling, and guarantee machine safety. (Lee et al. 2014)
- Automated analytics, semantics-based information fusion and process automation are among the targets for improving the performance of systems for real-time business intelligence (RTBI). Technologies like intelligent data analysis, soft computing and ontologies will play a major role in the development of RTBI. (Azvine et al. 2006)
- Further research must be directed towards dynamic optimization of production systems and in combination with resource allocation systems, so that manufacturing resources can be used in an optimal way (Zhang et al. 2018).

4.2 Utilization of artificial or computational intelligence

This category underlines the importance of AI/CI techniques and models to provide solutions to complicated, real-world problems, such as

- incorporating pattern recognition methods based on similarity measures to enhance queries by relaxing the matching functions using the semantic distances between the qualitative episodes (Roda and Musulin 2014)
- improving the used approach with stream reasoning capabilities. Stream reasoning is a subject of topical interest for the Semantic Web that aims at providing high-level skills for processing time stamped data (Roda and Musulin 2014)
- developing more advanced analytical methods to handle the data deluge, such as topic modeling and deep learning (Luckow et al. 2015)
- producing scalable machine learning approaches that are essential to extract knowledge from data (Luckow et al. 2015)
- designing and implementing intelligent, data-driven services (Luckow et al. 2015)
- investigating support for a wide range of machine learning tasks and for a more asynchronous, GraphLab-inspired programming models for encoding graphical algorithms (Borkar et al. 2012)
- delivering advanced analytic techniques, such as deep machine learning algorithm, that will allow computers to detect items of interest in large quantities of unstructured data, and to deduce relationships without needing specific models or programming instructions (Cuenca and Mehdi 2011; Mehdi et al. 2017; Wang et al. 2018)

4.3 Enhancement of experimentation

The majority of the papers reviewed described the necessity to enhance and extend their experimental part towards assessing generalizability, scalability, performance and integrate-ability of their models and approaches. For example, Cuenca and Mehdi (<year>), Mehdi et al. (2017) and Wang et al. (2018) address the issue of Maturity of the proposed models, from prototypes to stable systems, and their real-world applicability. They describe the need for more extensive scalability experiments, more intensive evaluation and performance comparison with other data-driven solutions, and argue about productivity and the ability to generate insight from data. They also suggest the use of hybrid query engines and the support for analytics across data residing on different platforms. Finally, they consider the maturity of available platforms and tools needed to meet the requirements of the increasing number of applications and users.

Larson and Chang (2016) suggests adopting business intelligence platforms, applications and services for all types of organizations. Roh, Heo, and Whang (2019) address the issues of Data

Evaluation, Performance Trade-off, Crowd-sourcing, as well as the empirical comparison of techniques, Generalizing and integrating techniques.

In addition to requiring empirical analysis of big data analytics enabled transformation, the work in Wang et al. (2018) also exposes the need for more scientific and quantitative studies, focusing on some of the business analytics capability elements identified.

4.4 IoT infrastructure, technology and applications

This category lists challenges that involve problems in IoT infrastructure and technology in general, as well as corresponding applications:

Lee and Lee (2015) suggests that before IoT id widely adopted by enterprises data centers must face challenges related to security, the enterprise, consumer privacy, data itself, storage management, server technologies, and networking. Mahdavinejad et al. (2018) moves along the same lines focusing IoT data characteristics, IoT applications and IoT data analytic algorithms. Bashir and Gill (2017) argue that there is a growing interest in IoT-enabled smart buildings, however, the storage and analysis of large amount of high-speed real-time smart building data is a challenging task. There are a number of contemporary Big Data management technologies and advanced analytics techniques that can be used to deal with this challenge. Tao et al. (2014) report that many challenges remain to be addressed before their proposal for the CCIoT-CMfg is implemented and applied, such as design and manufacturing of high-frequency chip antenna, special sensors, as well as the deployment technologies, e.g., optical fiber sensors for online and real-time monitoring high-speed rotating equipment with high working temperature. The authors also state that the majority of the papers that deal with IoT in manufacturing focus on the data collection of manufacturing equipment and process, but the corresponding literature on how to realize the intelligent data mining and processing of these collected data, and generate the useful information to serve the manufacturing requirement is insufficient, lacking standardization, protocols, safety, reliability, and management level of applying IoT and CC in manufacturing. In addition, there is contradiction between manufacturing resource/information sharing and protection of privacy/core technology by using IoT and CC technologies in manufacturing.

4.5 Data modeling and characteristics

Papers in this broad category report challenges pertaining data modelling, data characteristics and data management:

Petersen et al. (2017) report that ontologies dedicated to different industry domains need to be developed to enable data integration and semantic interoperability within and between companies in conjunction with related business processes and governance models. Also, further support for ontology-based data access would be needed to achieve the envisioned scalability.

Lee and Lee (2015) consider data management, data mining, privacy and security as the most important challenges.

The work of Qin et al. (2016) goes a step further and reports the following as open research problems of this area:

- Data quality and uncertainty
- Co-space data
- Transaction handling
- Frequently Updated Timestamped Structured (FUTS) data
- Distributed and mobile data
- Semantic enrichment and semantic event processing
- Mining
- Knowledge discovery
- Security
- Privacy
- Social concerns

Khine and Wang (2018) deal with data lakes challenges, and describes the following challenges:

- Data lakes lack the ability to determine data quality or the lineage of findings. Other data analysts have found out them in the same data lake but cannot provide for later analysts.
- Data Lakes accept any data without oversight and governance.
- There is no descriptive metadata or a mechanism to maintain metadata leading to data swamp.
- Data need to analyze from scratch every time.
- Performance cannot be guaranteed.
- Security (privacy and regulatory requirements) and access control (weakness of metadata management) as data in a lake can be replaced without oversight of the contents.

The authors in Chungoora et al. (2013) state that Manufacturing Ontology should be extended into a reference ontology model for capturing product lifecycle knowledge with increasing complexity. Such an ontology should be able to accommodate concepts central to, e.g., machine control levels and discrete manufacturing timescales, operational timescales needed for inter-machine configurations, longer term product configurations and product service-systems

4.6 Cloud environment

This category includes papers that report challenges revolving around the Cloud and smart data processing for control, configuration and monitoring.

Aceto et al. (2013) put forward a set of challenges that Cloud monitoring systems will have to face in the future, such as, effectiveness, efficiency, new monitoring techniques and tools, cross-layer monitoring, cross-domain monitoring: federated clouds, hybrid clouds, multi-tenancy services, and monitoring of novel network architectures

Saldivar et al. (2016) suggest future directions for a higher scalable application on predictive way to select attributes lead to focus on powerful tools like fuzzy logic for fuzzified mass customization.

Wang et al. (2018) argue that as the evolution of computing resources (e.g., cloud computing, fog computing, etc.), computational intelligence including deep learning may be pushed into Cloud, enabling more convenient and on-demand computing services for smart manufacturing.

4.7 Smart data processing

A variety of research challenges are reported in this category pertaining the production and processing of smart data:

Denno et al. (2018) report the need to use the models in real-time production-control decision making and to explore integrated methodologies with smart manufacturing operational technology.

Papazoglou et al. (2015) describe that smart manufacturing combines technology, knowledge, information, and humans in manufacturing intelligence to every aspect of applications and that it fundamentally changes how products are manufactured and delivered. In this context the authors expect that product innovations will arise from the creative use of manufacturing knowledge gathered from every point of an SMN value chain, ranging from consumer preferences to production and delivery mechanisms.

Lu and Xu (2019) argue that real-time machine control over the Internet is considered a very challenging task and a critical milestone to enable distributed smart factories and smart manufacturing. The authors also state that the existing TCP/UDP/WebSocket protocol is not suitable for CNC machine interpolated data to be transmitted through WAN connection. Strategies need to be developed to compensate the effect resulting from unpredictable WAN connection.

The authors in Petersen et al. (2016) propose a number of directions for future work. In particular, the Semantic Factory approach could be expanded from single factories to an integration approach covering the entire enterprise, as well as supply networks. Also, the exploitation of the integrated data for advanced analytics and forecasting is a promising area.

Mehdi et al. (2017) describe that in the future, CPS configuration and operation theory for intelligent manufacturing in shop floor must be further studied based on the proposed CPS architecture. Also, that the evaluation of dynamic manufacturing capability in shop floor should be conducted by processing the condition monitoring data of machining equipment.

The future direction of Papazoglou and Elgammal (2018) targets the concept of "self-organizing manufacturing processes for highly customizable products". It is expected that the blueprinting approach will evolve towards autonomic, reconfigurable manufacturing systems where a manufacturer receives a digital blueprint model of a new product, and based on the information in the model, the production environment will configure itself to produce that product. Autonomic, re-configurable manufacturing takes all the manufacturing to the next level by combining flexibility and self-adaptability of the production systems, self-optimization of adjustable smart production resources across all functions, paving the way for new product and more agile service platform schemes.

Finally, Lin et al. (2012) focus on future steps needed to model other decision-making processes in compliance with the CDSM for the core BPs. Example of these BPs for sustainability in manufacturing include collaborative product design where design objectives of individual functional units are considered at the CMN level, and production order allocation problem where integer-based Meta-Goal Programming process is used to optimally distribute production orders amongst a group of manufacturers. Thus, the GDSS establishes a fundamental platform for future applications of decision-support approaches that can be continuously added in supporting the dynamic management processes for achieving sustainability in manufacturing in a CMN.

4.8 Standardization

This category includes papers that draw attention to standardization issues and open problems in all areas of research of this survey.

Pang et al. (2015) describe firstly that data source agents and wrapper agents must be extended to support more types of data sources through accommodating more drivers and API library, and secondly, the extension of the data model library to support more international standards in logistic and manufacturing domain. The authors argue that more efforts are required to better understand these standards for better exploitation. Finally, ontology learn mechanism must be devised to enable the ontology self-update its library in order to improve the results.

Harjunkoski and Bauer (2014) suggest that a natural next step in research is to develop algorithmic libraries and methods that accept ISA-95-based instances as an input. This could pave the way toward a much sought-after holistic scheduling solver that analyzes the problem and selects or recommends the most suitable method for solving it. Having a solver that can be used by modelers without algorithmic knowledge would mean a tremendous boost in the industrial applicability of scheduling solutions.

Chungoora et al. (2013) suggest that parallel progress should also be made to pursue the semantic consolidation of concepts across model-based standards in industrial automation. Furthermore, in order to apply the model-driven concept within other spheres of knowledge, the authors conclude that the notion of foundation ontologies will need to be exploited so as to leverage interoperability.

Finally, Petersen et al. (2017) propose the continuous translation of relevant industry concepts and standards into RDF, as well as their integration and alignment with existing ontologies and vocabularies.

5. Conclusions

This survey was conducted in the context of the DESTINI project aiming at finding and studying articles published in the general area of Smart Data so as, on one hand, to form a solid scientific background that will enable its consortium to devise a research agenda for investigating selected topics in this area, and on the other, to identify the most significant challenges and open problems. The latter will form the basis for selecting specific problems that DESTINI will focus on to provide solutions or add a significant piece in the puzzle of understanding their dynamics and complexity.

In this context 81 papers were selected and studied, which fall in one or more of the three areas identified in DESTINI to be the main research pillars: (i) Smart Data Processing Systems, (ii) Systems of Deep Insight, and, (iii) Methodology for Data-Centric Services and Applications. The methodology used to search and locate these papers was the one usually employed for conducting systematic surveys, while the findings reported summarize the latest and most significant advances in the aforementioned pillars. This survey was completed by listing briefly the most challenging open problems described in the papers reviewed, organised in eight categories: (i) Optimization, Decision Support, Prediction; (ii) Utilization of artificial or computational intelligence; (iii) Enhancement of experimentation; (iv) IoT infrastructure, technology and applications; (v) Data modeling and characteristics; (vi) Cloud environment; (vii) Smart data processing; and, (viii) Standardization. These challenges will feed deliverable D2.3: Research & Innovation Agenda which will outline the research subjects of interest and devise a roadmap for their investigation.

References

Aceto, Giuseppe, Alessio Botta, Walter De Donato, and Antonio Pescapè. 2013. "Cloud Monitoring: A Survey." *Computer Networks* 57(9): 2093–2115. http://dx.doi.org/10.1016/j.comnet.2013.04.001.

Azvine, B., Z. Cui, D. D. Nauck, and B. Majeed. 2006. "Real Time Business Intelligence for the Adaptive Enterprise." *CEC/EEE 2006 Joint Conferences* 2006.

Azvine, Ben, Zhan Cui, and Detlef D. Nauck. 2005. "Towards Real-Time Business Intelligence." *BT Technology Journal* 23(3): 214–25.

Barnaghi, Payam, Amit Sheth, and Cory Henson. 2013. "From Data to Actionable Knowledge: Big Data Challenges in the Web of Things." *IEEE Intelligent Systems* 28(6).

Bashir, Muhammad Rizwan, and Asif Qumer Gill. 2017. "Towards an IoT Big Data Analytics Framework: Smart Buildings Systems." *Proceedings - 18th IEEE International Conference on High Performance Computing and Communications, 14th IEEE International Conference on Smart City and 2nd IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2016*: 1325–32.

Bertino, Elisa. 2013. "Big Data - Opportunities and Challenges: Panel Position Paper." *Proceedings - International Computer Software and Applications Conference*: 479–80.

Blazquez, Desamparados, and Josep Domenech. 2018. "Big Data Sources and Methods for Social and Economic Analyses." *Technological Forecasting and Social Change* 130(March 2017): 99–113. https://doi.org/10.1016/j.techfore.2017.07.027.

Bock, Conrad, Xuanfang Zha, Hyo Won Suh, and Jae Hyun Lee. 2010. "Ontological Product Modeling for Collaborative Design." *Advanced Engineering Informatics* 24(4): 510–24.

Borkar, V et al. 2012. "Declarative Systems for Large-Scale Machine Learning." *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*: 1–9. ftp://ftp.research.microsoft.com/pub/debull/A12june/declare.pdf.

Cafarella, Michael J., Alon Halevy, and Nodira Khoussainova. 2009. "Data Integration for the Relational Web." *Proceedings of the VLDB Endowment* 2(1): 1090–1101.

Charest, Michel, and Sylvain Delisle. 2006. "Ontology-Guided Intelligent Data Mining Assistance: Combining Declarative and Procedural Knowledge." *Proceedings of the 10th IASTED International Conference on Artificial Intelligence and Soft Computing, ASC 2006*: 9–14.

Chen, Hsinchun, Roger H. L. Chiang, and Veda C. Storey. 2012. "B Usiness I Ntelligence and a Nalytics." *MIS Quarterly* 36(4): 1165–88.

Chen, Min, Shiwen Mao, and Yunhao Liu. 2014. "Big Data: A Survey." *Mobile Networks and Applications* 19(2): 171–209.

Chungoora, Nitishal et al. 2013. "A Model-Driven Ontology Approach for Manufacturing System Interoperability and Knowledge Sharing." *Computers in Industry* 64(4): 392–401.

Cuenca, E Kharlamov B, Grau E Jim, and S Lamparter G Mehdi. "Capturing Industrial Information Models with Ontologies and Constraints : The Siemens Use Case." 2.

Dean, Jeffrey, and Sanjay Ghemawat. 2008. "MapReduce: Simplified Data Processing on Large Clusters." *Communications of the ACM* 51(1): 107–13.

Delen, Dursun, and Haluk Demirkan. 2013. "Data, Information and Analytics as Services." *Decision Support Systems* 55(1): 359–63. http://dx.doi.org/10.1016/j.dss.2012.05.044.

Denno, Peter, Charles Dickerson, and Jennifer Anne Harding. 2018. "Dynamic Production System Identification for Smart Manufacturing Systems." *Journal of Manufacturing Systems* 48(May): 192–203. https://doi.org/10.1016/j.jmsy.2018.04.006.

Dong, Xin Luna, Barna Saha, and Divesh Srivastava. 2012. "Less Is More: Selecting Sources

Wisely for Integration.” *Proceedings of the VLDB Endowment* 6(2): 37–48.

Drabent, Włodzimierz et al. 2009. “Hybrid Reasoning with Rules and Ontologies.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5500 LNCS: 1–49.

Erkin, Zekeriya, Juan Ramón Troncoso-Pastoriza, R. L. Lagendijk, and Fernando Pérez-González. 2013. “Privacy-Preserving Data Aggregation in Smart Metering Systems: An Overview.” *IEEE Signal Processing Magazine* 30(2): 75–86.

Fang, Huang. 2015. “Managing Data Lakes in Big Data Era: What’s a Data Lake and Why Has It Became Popular in Data Management Ecosystem.” *2015 IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, IEEE-CYBER 2015*: 820–24.

Farid, Mina et al. 2016. “CLAMS: Bringing Quality to Data Lakes.” *Proceedings of the ACM SIGMOD International Conference on Management of Data* 26-June-20: 2089–92.

Garetti, M., L. Fumagalli, A. Lobov, and J. L. Martinez Lastra. 2013. 46 IFAC Proceedings Volumes (IFAC-PapersOnline) *Open Automation of Manufacturing Systems through Integration of Ontology and Web Services*. IFAC. http://dx.doi.org/10.3182/20130619-3-RU-3018.00169.

Gharaibeh, Ammar et al. 2017. “Smart Cities: A Survey on Data Management, Security, and Enabling Technologies.” *IEEE Communications Surveys and Tutorials* 19(4): 2456–2501.

Giese, Martin et al. 2015. “Optique – Zooming In on Big Data Access.” *Computer*.

Giovannini, Antonio et al. 2012. “Ontology-Based System for Supporting Manufacturing Sustainability.” *Annual Reviews in Control* 36(2): 309–17. http://dx.doi.org/10.1016/j.arcontrol.2012.09.012.

Guerrero, Juan I. et al. 2017. “Heterogeneous Data Source Integration for Smart Grid Ecosystems Based on Metadata Mining.” *Expert Systems with Applications* 79: 254–68. http://dx.doi.org/10.1016/j.eswa.2017.03.007.

Günther, Wendy Arianne, Mohammad H. Rezazade Mehrizi, Marleen Huysman, and Frans Feldberg. 2017. “Debating Big Data: A Literature Review on Realizing Value from Big Data.” *Journal of Strategic Information Systems* 26(3): 191–209.

Harjunkoski, Iiro, and Reinhard Bauer. 2014. “Sharing Data for Production Scheduling Using the ISA-95 Standard.” *Frontiers in Energy Research* 2(OCT).

Jardim-Goncalves, Ricardo et al. 2014. “Collaborative Negotiation for Ontology-Driven Enterprise Businesses.” *Computers in Industry* 65(9): 1232–41. http://dx.doi.org/10.1016/j.compind.2014.01.001.

Khine, Pwint Phyu, and Zhao Shun Wang. 2018. “Data Lake: A New Ideology in Big Data Era.” *ITM Web of Conferences* 17: 03025.

Kim, Junwhan. 2017. “Partial Rollback-Based Scheduling on In-Memory Transactional Data Grids.” *Big Data Research* 9: 47–56. http://dx.doi.org/10.1016/j.bdr.2017.06.004.

Kitchenham, Barbara et al. 2010. “Systematic Literature Reviews in Software Engineering-A Tertiary Study.” *Information and Software Technology* 52(8): 792–805. http://dx.doi.org/10.1016/j.infsof.2010.03.006.

Lanzenberger, Monika, Jennifer Sampson, and Markus Rester. 2010. “Ontology Visualization: Tools and Techniques for Visual Representation of Semi-Structured Meta-Data.” *Journal of Universal Computer Science* 16(7): 1036–54.

Larson, Deanne, and Victor Chang. 2016. “A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science.” *International Journal of Information Management* 36(5): 700–710. http://dx.doi.org/10.1016/j.ijinfomgt.2016.04.013.

Lee, In, and Kyoochun Lee. 2015. “The Internet of Things (IoT): Applications, Investments, and Challenges for Enterprises.” *Business Horizons* 58(4): 431–40.

http://dx.doi.org/10.1016/j.bushor.2015.03.008.

Lee, Jay, Hossein Davari Ardakani, Shanhu Yang, and Behrad Bagheri. 2015. "Industrial Big Data Analytics and Cyber-Physical Systems for Future Maintenance & Service Innovation." *Procedia CIRP* 38: 3–7. http://dx.doi.org/10.1016/j.procir.2015.08.026.

Lee, Jay, Behrad Bagheri, and Hung An Kao. 2015. "A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems." *Manufacturing Letters* 3: 18–23. http://dx.doi.org/10.1016/j.mfglet.2014.12.001.

Lee, Jay, Hung An Kao, and Shanhu Yang. 2014. "Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment." *Procedia CIRP* 16: 3–8. http://dx.doi.org/10.1016/j.procir.2014.02.001.

Lee, Jay, Edzel Lapira, Shanhu Yang, and Ann Kao. 2013. 46 IFAC Proceedings Volumes (IFAC-PapersOnline) *Predictive Manufacturing System - Trends of next-Generation Production Systems*. IFAC. http://dx.doi.org/10.3182/20130522-3-BR-4036.00107.

Lemaignan, Séverin, Ali Siadat, Jean Yves Dantan, and Anatoli Semenenko. 2006. "MASON: A Proposal for an Ontology of Manufacturing Domain." *Proceedings - DIS 2006: IEEE Workshop on Distributed Intelligent Systems - Collective Intelligence and Its Applications* 2006: 195–200.

Li, B. M., S. Q. Xie, and X. Xu. 2011. "Recent Development of Knowledge-Based Systems, Methods and Tools for One-of-a-Kind Production." *Knowledge-Based Systems* 24(7): 1108–19.

Lin, Hao W., Sev V. Nagalingam, Swee S. Kuik, and Tomohiro Murata. 2012. "Design of a Global Decision Support System for a Manufacturing SME: Towards Participating in Collaborative Manufacturing." *International Journal of Production Economics* 136(1): 1–12. http://dx.doi.org/10.1016/j.ijpe.2011.07.001.

Lu, Yuqian, and Xun Xu. 2019. "Cloud-Based Manufacturing Equipment and Big Data Analytics to Enable on-Demand Manufacturing Services." *Robotics and Computer-Integrated Manufacturing* 57(October 2018): 92–102. https://doi.org/10.1016/j.rcim.2018.11.006.

Lu, Zongqing, and Yonggang Wen. 2014. "Distributed Algorithm for Tree-Structured Data Aggregation Service Placement in Smart Grid." *IEEE Systems Journal* 8(2): 553–61.

Luckow, Andre et al. 2015. "Automotive Big Data: Applications, Workloads and Infrastructures." *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*: 1201–10.

Mahdavinejad, Mohammad Saeid et al. 2018. "Machine Learning for Internet of Things Data Analysis: A Survey." *Digital Communications and Networks* 4(3): 161–75. https://doi.org/10.1016/j.dcan.2017.10.002.

Mehdi, Gulnar et al. 2017. "Semantic Rule-Based Equipment Diagnostics." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10588 LNCS(3): 314–33. http://dx.doi.org/10.1016/j.procir.2014.02.001.

Miloslavskaya, Natalia, and Alexander Tolstoy. 2016. "Big Data, Fast Data and Data Lake Concepts." *Procedia Computer Science* 88: 300–305. http://dx.doi.org/10.1016/j.procs.2016.07.439.

O'Donovan, P., K. Leahy, K. Bruton, and D. T.J. O'Sullivan. 2015. "An Industrial Big Data Pipeline for Data-Driven Analytics Maintenance Applications in Large-Scale Smart Manufacturing Facilities." *Journal of Big Data* 2(1): 1–26.

O'Leary, Daniel E. 2014. "Embedding AI and Crowdsourcing in the Big Data Lake." *IEEE Intelligent Systems* 29(5): 70–73.

Pang, L. Y., Ray Y. Zhong, J. Fang, and George Q. Huang. 2015. "Data-Source Interoperability Service for Heterogeneous Information Integration in Ubiquitous Enterprises."

Advanced Engineering Informatics 29(3): 549–61. http://dx.doi.org/10.1016/j.aei.2015.04.007.

Papazoglou, Michael P., and Amal Elgammal. 2018. “The Manufacturing Blueprint Environment: Bringing Intelligence into Manufacturing.” *2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings* 2018-Janua: 750–59.

Papazoglou, Michael P., Willem Jan Van Den Heuvel, and Julien Etienne Mascolo. 2015. “A Reference Architecture and Knowledge-Based Structures for Smart Manufacturing Networks.” *IEEE Software* 32(3): 61–69.

Petersen, Niklas et al. 2016. “Monitoring and Automating Factories Using Semantic Models.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10055 LNCS: 315–30.

———. 2017. “Realizing an RDF-Based Information Model for, a Manufacturing Company – A Case Study.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10588 LNCS: 350–66.

Polyvyanyy, Artem, Chun Ouyang, Alistair Barros, and Wil M.P. van der Aalst. 2017. “Process Querying: Enabling Business Intelligence through Query-Based Process Analytics.” *Decision Support Systems* 100: 41–56. http://dx.doi.org/10.1016/j.dss.2017.04.011.

Professor, Associate. 2008. “Product Configuration and Product Family Modelling.” : 1–21.

Qin, Yongrui et al. 2016. “When Things Matter: A Survey on Data-Centric Internet of Things.” *Journal of Network and Computer Applications* 64: 137–53. http://dx.doi.org/10.1016/j.jnca.2015.12.016.

Roda, Fernando, and Estanislao Musulin. 2014. “An Ontology-Based Framework to Support Intelligent Data Analysis of Sensor Measurements.” *Expert Systems with Applications* 41(17): 7914–26. http://dx.doi.org/10.1016/j.eswa.2014.06.033.

Roh, Yuji, Geon Heo, and Steven Euijong Whang. 2019. “A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective.” *IEEE Transactions on Knowledge and Data Engineering* PP(c): 1–1.

Rusitschka, Sebnem, Kolja Eger, and Christoph Gerdes. 2010. “Smart Grid Data Cloud: A Model for Utilizing Cloud Computing in the Smart Grid Domain.” *2010 First IEEE International Conference on Smart Grid Communications*: 483–88.

Sahay, B. S., and Jayanthi Ranjan. 2008. “Real Time Business Intelligence in Supply Chain Analytics.” *Information Management and Computer Security* 16(1): 28–48.

Sakr, Sherif, and Amal Elgammal. 2016. “Towards a Comprehensive Data Analytics Framework for Smart Healthcare Services.” *Big Data Research* 4: 44–58. http://dx.doi.org/10.1016/j.bdr.2016.05.002.

Saldivar, Alfredo Alan Flores et al. 2016. “Identifying Smart Design Attributes for Industry 4.0 Customization Using a Clustering Genetic Algorithm.” *2016 22nd International Conference on Automation and Computing, ICAC 2016: Tackling the New Challenges in Automation and Computing* (September): 408–14.

Sethi, Pallavi, and Smruti R. Sarangi. 2017. “Internet of Things: Architectures, Protocols, and Applications.” *Journal of Electrical and Computer Engineering* 2017.

Stimmel, Carol L. 2016. Big Data Analytics Strategies for the Smart Grid *Big Data Analytics Strategies for the Smart Grid*.

Stojmenovic, Ivan. 2014. “Machine-to-Machine Communications with in-Network Data Aggregation, Processing, and Actuation for Large-Scale Cyber-Physical Systems.” *IEEE Internet of Things Journal* 1(2): 122–28.

Sun, Jianing et al. 2015. “Smart Services for Enhancing Personal Competence in Industrie 4.0 Digital Factory.” *Logforum* 14(1): 51–57.

Tang, Dunbing et al. 2018. “Using Autonomous Intelligence to Build a Smart Shop Floor.” *International Journal of Advanced Manufacturing Technology* 94(5–8): 1597–1606.

Tao, Fei et al. 2014. “CCIoT-CMfg: Cloud Computing and Internet of Things-Based Cloud Manufacturing Service System.” *IEEE Transactions on Industrial Informatics* 10(2): 1435–42.

Terkaj, W., and M. Urgo. 2015. “A Virtual Factory Data Model as a Support Tool for the Simulation of Manufacturing Systems.” *Procedia CIRP* 28: 137–42. http://dx.doi.org/10.1016/j.procir.2015.04.023.

Wang, Jinjiang et al. 2018. “Deep Learning for Smart Manufacturing: Methods and Applications.” *Journal of Manufacturing Systems* 48: 144–56. https://doi.org/10.1016/j.jmsy.2018.01.003.

Wang, Jinjiang, Peilun Fu, and Robert X. Gao. 2019. “Machine Vision Intelligence for Product Defect Inspection Based on Deep Learning and Hough Transform.” *Journal of Manufacturing Systems* 51(March): 52–60. https://doi.org/10.1016/j.jmsy.2019.03.002.

Wang, Shiyong et al. 2016. “Towards Smart Factory for Industry 4.0: A Self-Organized Multi-Agent System with Big Data Based Feedback and Coordination.” *Computer Networks* 101: 158–68.

Wang, Yichuan, Lee Ann Kung, and Terry Anthony Byrd. 2018. “Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations.” *Technological Forecasting and Social Change* 126: 3–13. https://doi.org/10.1016/j.techfore.2015.12.019.

Wohlin, Claes. 2014. “Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering.” *ACM International Conference Proceeding Series*.

Yan, Ye, Yi Qian, and Hamid Sharif. 2011. “A Secure Data Aggregation and Dispatch Scheme for Home Area Networks in Smart Grid.” *GLOBECOM - IEEE Global Telecommunications Conference*: 1–6.

Yang, Dong, Ming Dong, and Rui Miao. 2008. “Development of a Product Configuration System with an Ontology-Based Approach.” *CAD Computer Aided Design* 40(8): 863–78.

Yu, Cui, and Josef Boyd. 2016. “FB+-Tree for Big Data Management.” *Big Data Research* 4: 25–36. http://dx.doi.org/10.1016/j.bdr.2015.11.003.

Zhang, Yan et al. 2012. “Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid.” *IEEE Network* 26(3): 6–13.

Zhang, Yingfeng et al. 2018. “Coloured Petri Net-Based Active Sensing System of Real-Time and Multi-Source Manufacturing Information for Smart Factory.” *International Journal of Advanced Manufacturing Technology* 94(9–12): 3427–39.

Zhong, Ray Y. et al. 2015. “A Big Data Approach for Logistics Trajectory Discovery from RFID-Enabled Production Data.” *International Journal of Production Economics* 165: 260–72.

Aceto, Giuseppe, Alessio Botta, Walter De Donato, and Antonio Pescapè. 2013. “Cloud Monitoring: A Survey.” *Computer Networks* 57(9): 2093–2115. http://dx.doi.org/10.1016/j.comnet.2013.04.001.

Azvine, B., Z. Cui, D. D. Nauck, and B. Majeed. 2006. “Real Time Business Intelligence for the Adaptive Enterprise.” *CEC/EEE 2006 Joint Conferences* 2006.

Azvine, Ben, Zhan Cui, and Detlef D. Nauck. 2005. “Towards Real-Time Business Intelligence.” *BT Technology Journal* 23(3): 214–25.

Barnaghi, Payam, Amit Sheth, and Cory Henson. 2013. “From Data to Actionable Knowledge: Big Data Challenges in the Web of Things.” *IEEE Intelligent Systems* 28(6).

Bashir, Muhammad Rizwan, and Asif Qumer Gill. 2017. “Towards an IoT Big Data Analytics Framework: Smart Buildings Systems.” *Proceedings - 18th IEEE International Conference on High Performance Computing and Communications, 14th IEEE International Conference on Smart City and 2nd IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2016*: 1325–32.

Bertino, Elisa. 2013. “Big Data - Opportunities and Challenges: Panel Position Paper.” *Proceedings - International Computer Software and Applications Conference*: 479–80.

Blazquez, Desamparados, and Josep Domenech. 2018. “Big Data Sources and Methods for Social and Economic Analyses.” *Technological Forecasting and Social Change* 130(March 2017): 99–113. https://doi.org/10.1016/j.techfore.2017.07.027.

Bock, Conrad, Xuanfang Zha, Hyo Won Suh, and Jae Hyun Lee. 2010. “Ontological Product Modeling for Collaborative Design.” *Advanced Engineering Informatics* 24(4): 510–24.

Borkar, V et al. 2012. “Declarative Systems for Large-Scale Machine Learning.” *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*: 1–9. ftp://ftp.research.microsoft.com/pub/debull/A12june/declare.pdf.

Cafarella, Michael J., Alon Halevy, and Nodira Khoussainova. 2009. “Data Integration for the Relational Web.” *Proceedings of the VLDB Endowment* 2(1): 1090–1101.

Charest, Michel, and Sylvain Delisle. 2006. “Ontology-Guided Intelligent Data Mining Assistance: Combining Declarative and Procedural Knowledge.” *Proceedings of the 10th IASTED International Conference on Artificial Intelligence and Soft Computing, ASC 2006*: 9–14.

Chen, Hsinchun, Roger H. L. Chiang, and Veda C. Storey. 2012. “B Usiness I Ntelligence and a Nalytics.” *MIS Quarterly* 36(4): 1165–88.

Chen, Min, Shiwen Mao, and Yunhao Liu. 2014. “Big Data: A Survey.” *Mobile Networks and Applications* 19(2): 171–209.

Chungoora, Nitishal et al. 2013. “A Model-Driven Ontology Approach for Manufacturing System Interoperability and Knowledge Sharing.” *Computers in Industry* 64(4): 392–401.

Cuenca, E Kharlamov B, Grau E Jim, and S Lamparter G Mehdi. “Capturing Industrial Information Models with Ontologies and Constraints : The Siemens Use Case.” 2.

Dean, Jeffrey, and Sanjay Ghemawat. 2008. “MapReduce: Simplified Data Processing on Large Clusters.” *Communications of the ACM* 51(1): 107–13.

Delen, Dursun, and Haluk Demirkan. 2013. “Data, Information and Analytics as

Services.” *Decision Support Systems* 55(1): 359–63. http://dx.doi.org/10.1016/j.dss.2012.05.044.

Denno, Peter, Charles Dickerson, and Jennifer Anne Harding. 2018. “Dynamic Production System Identification for Smart Manufacturing Systems.” *Journal of Manufacturing Systems* 48(May): 192–203. https://doi.org/10.1016/j.jmsy.2018.04.006.

Dong, Xin Luna, Barna Saha, and Divesh Srivastava. 2012. “Less Is More: Selecting Sources Wisely for Integration.” *Proceedings of the VLDB Endowment* 6(2): 37–48.

Drabent, Włodzimierz et al. 2009. “Hybrid Reasoning with Rules and Ontologies.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5500 LNCS: 1–49.

Erkin, Zekeriya, Juan Ramón Troncoso-Pastoriza, R. L. Lagendijk, and Fernando Pérez-González. 2013. “Privacy-Preserving Data Aggregation in Smart Metering Systems: An Overview.” *IEEE Signal Processing Magazine* 30(2): 75–86.

Fang, Huang. 2015. “Managing Data Lakes in Big Data Era: What’s a Data Lake and Why Has It Became Popular in Data Management Ecosystem.” *2015 IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, IEEE-CYBER 2015*: 820–24.

Farid, Mina et al. 2016. “CLAMS: Bringing Quality to Data Lakes.” *Proceedings of the ACM SIGMOD International Conference on Management of Data* 26-June-20: 2089–92.

Garetti, M., L. Fumagalli, A. Lobov, and J. L. Martinez Lastra. 2013. 46 IFAC Proceedings Volumes (IFAC-PapersOnline) *Open Automation of Manufacturing Systems through Integration of Ontology and Web Services*. IFAC. http://dx.doi.org/10.3182/20130619-3-RU-3018.00169.

Gharaibeh, Ammar et al. 2017. “Smart Cities: A Survey on Data Management, Security, and Enabling Technologies.” *IEEE Communications Surveys and Tutorials* 19(4): 2456–2501.

Giese, Martin et al. 2015. “Optique – Zooming In on Big Data Access.” *Computer*.

Giovannini, Antonio et al. 2012. “Ontology-Based System for Supporting Manufacturing Sustainability.” *Annual Reviews in Control* 36(2): 309–17. http://dx.doi.org/10.1016/j.arcontrol.2012.09.012.

Guerrero, Juan I. et al. 2017. “Heterogeneous Data Source Integration for Smart Grid Ecosystems Based on Metadata Mining.” *Expert Systems with Applications* 79: 254–68. http://dx.doi.org/10.1016/j.eswa.2017.03.007.

Günther, Wendy Arianne, Mohammad H. Rezazade Mehrizi, Marleen Huysman, and Frans Feldberg. 2017. “Debating Big Data: A Literature Review on Realizing Value from Big Data.” *Journal of Strategic Information Systems* 26(3): 191–209.

Harjunkoski, Iiro, and Reinhard Bauer. 2014. “Sharing Data for Production Scheduling Using the ISA-95 Standard.” *Frontiers in Energy Research* 2(OCT).

Jardim-Goncalves, Ricardo et al. 2014. "Collaborative Negotiation for Ontology-Driven Enterprise Businesses." *Computers in Industry* 65(9): 1232–41. http://dx.doi.org/10.1016/j.compind.2014.01.001.

Khine, Pwint Phyu, and Zhao Shun Wang. 2018. "Data Lake: A New Ideology in Big Data Era." *ITM Web of Conferences* 17: 03025.

Kim, Junwhan. 2017. "Partial Rollback-Based Scheduling on In-Memory Transactional Data Grids." *Big Data Research* 9: 47–56. http://dx.doi.org/10.1016/j.bdr.2017.06.004.

Kitchenham, Barbara et al. 2010. "Systematic Literature Reviews in Software Engineering-A Tertiary Study." *Information and Software Technology* 52(8): 792–805. http://dx.doi.org/10.1016/j.infsof.2010.03.006.

Lanzenberger, Monika, Jennifer Sampson, and Markus Rester. 2010. "Ontology Visualization: Tools and Techniques for Visual Representation of Semi-Structured Meta-Data." *Journal of Universal Computer Science* 16(7): 1036–54.

Larson, Deanne, and Victor Chang. 2016. "A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science." *International Journal of Information Management* 36(5): 700–710. http://dx.doi.org/10.1016/j.ijinfomgt.2016.04.013.

Lee, In, and Kyoochun Lee. 2015. "The Internet of Things (IoT): Applications, Investments, and Challenges for Enterprises." *Business Horizons* 58(4): 431–40. http://dx.doi.org/10.1016/j.bushor.2015.03.008.

Lee, Jay, Hossein Davari Ardakani, Shanhu Yang, and Behrad Bagheri. 2015. "Industrial Big Data Analytics and Cyber-Physical Systems for Future Maintenance & Service Innovation." *Procedia CIRP* 38: 3–7. http://dx.doi.org/10.1016/j.procir.2015.08.026.

Lee, Jay, Behrad Bagheri, and Hung An Kao. 2015. "A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems." *Manufacturing Letters* 3: 18–23. http://dx.doi.org/10.1016/j.mfglet.2014.12.001.

Lee, Jay, Hung An Kao, and Shanhu Yang. 2014. "Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment." *Procedia CIRP* 16: 3–8. http://dx.doi.org/10.1016/j.procir.2014.02.001.

Lee, Jay, Edzel Lapira, Shanhu Yang, and Ann Kao. 2013. 46 IFAC Proceedings Volumes (IFAC-PapersOnline) *Predictive Manufacturing System - Trends of next-Generation Production Systems*. IFAC. http://dx.doi.org/10.3182/20130522-3-BR-4036.00107.

Lemaignan, Séverin, Ali Siadat, Jean Yves Dantan, and Anatoli Semenenko. 2006. "MASON: A Proposal for an Ontology of Manufacturing Domain." *Proceedings - DIS 2006: IEEE Workshop on Distributed Intelligent Systems - Collective Intelligence and Its Applications* 2006: 195–200.

Li, B. M., S. Q. Xie, and X. Xu. 2011. "Recent Development of Knowledge-Based Systems, Methods and Tools for One-of-a-Kind Production." *Knowledge-Based Systems* 24(7): 1108–19.

Lin, Hao W., Sev V. Nagalingam, Swee S. Kuik, and Tomohiro Murata. 2012. "Design of a Global Decision Support System for a Manufacturing SME: Towards Participating in Collaborative Manufacturing." *International Journal of Production Economics* 136(1): 1–12. http://dx.doi.org/10.1016/j.ijpe.2011.07.001.

Lu, Yuqian, and Xun Xu. 2019. "Cloud-Based Manufacturing Equipment and Big Data Analytics to Enable on-Demand Manufacturing Services." *Robotics and Computer-Integrated Manufacturing* 57(October 2018): 92–102. https://doi.org/10.1016/j.rcim.2018.11.006.

Lu, Zongqing, and Yonggang Wen. 2014. "Distributed Algorithm for Tree-Structured Data Aggregation Service Placement in Smart Grid." *IEEE Systems Journal* 8(2): 553–61.

Luckow, Andre et al. 2015. "Automotive Big Data: Applications, Workloads and Infrastructures." *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*: 1201–10.

Mahdavinejad, Mohammad Saeid et al. 2018. "Machine Learning for Internet of Things Data Analysis: A Survey." *Digital Communications and Networks* 4(3): 161–75. https://doi.org/10.1016/j.dcan.2017.10.002.

Mehdi, Gulnar et al. 2017. "Semantic Rule-Based Equipment Diagnostics." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10588 LNCS(3): 314–33. http://dx.doi.org/10.1016/j.procir.2014.02.001.

Miloslavskaya, Natalia, and Alexander Tolstoy. 2016. "Big Data, Fast Data and Data Lake Concepts." *Procedia Computer Science* 88: 300–305. http://dx.doi.org/10.1016/j.procs.2016.07.439.

O'Donovan, P., K. Leahy, K. Bruton, and D. T.J. O'Sullivan. 2015. "An Industrial Big Data Pipeline for Data-Driven Analytics Maintenance Applications in Large-Scale Smart Manufacturing Facilities." *Journal of Big Data* 2(1): 1–26.

O'Leary, Daniel E. 2014. "Embedding AI and Crowdsourcing in the Big Data Lake." *IEEE Intelligent Systems* 29(5): 70–73.

Pang, L. Y., Ray Y. Zhong, J. Fang, and George Q. Huang. 2015. "Data-Source Interoperability Service for Heterogeneous Information Integration in Ubiquitous Enterprises." *Advanced Engineering Informatics* 29(3): 549–61. http://dx.doi.org/10.1016/j.aei.2015.04.007.

Papazoglou, Michael P., and Amal Elgammal. 2018. "The Manufacturing Blueprint Environment: Bringing Intelligence into Manufacturing." *2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings* 2018-Janua: 750–59.

Papazoglou, Michael P., Willem Jan Van Den Heuvel, and Julien Etienne Mascolo. 2015. "A Reference Architecture and Knowledge-Based Structures for Smart Manufacturing Networks." *IEEE Software* 32(3): 61–69.

Petersen, Niklas et al. 2016. “Monitoring and Automating Factories Using Semantic Models.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10055 LNCS: 315–30.

———. 2017. “Realizing an RDF-Based Information Model for, a Manufacturing Company – A Case Study.” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10588 LNCS: 350–66.

Polyvyanyy, Artem, Chun Ouyang, Alistair Barros, and Wil M.P. van der Aalst. 2017. “Process Querying: Enabling Business Intelligence through Query-Based Process Analytics.” *Decision Support Systems* 100: 41–56. http://dx.doi.org/10.1016/j.dss.2017.04.011.

Professor, Associate. 2008. “Product Configuration and Product Family Modelling.” : 1–21.

Qin, Yongrui et al. 2016. “When Things Matter: A Survey on Data-Centric Internet of Things.” *Journal of Network and Computer Applications* 64: 137–53. http://dx.doi.org/10.1016/j.jnca.2015.12.016.

Roda, Fernando, and Estanislao Musulin. 2014. “An Ontology-Based Framework to Support Intelligent Data Analysis of Sensor Measurements.” *Expert Systems with Applications* 41(17): 7914–26. http://dx.doi.org/10.1016/j.eswa.2014.06.033.

Roh, Yuji, Geon Heo, and Steven Euijong Whang. 2019. “A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective.” *IEEE Transactions on Knowledge and Data Engineering* PP(c): 1–1.

Rusitschka, Sebnem, Kolja Eger, and Christoph Gerdes. 2010. “Smart Grid Data Cloud: A Model for Utilizing Cloud Computing in the Smart Grid Domain.” *2010 First IEEE International Conference on Smart Grid Communications*: 483–88.

Sahay, B. S., and Jayanthi Ranjan. 2008. “Real Time Business Intelligence in Supply Chain Analytics.” *Information Management and Computer Security* 16(1): 28–48.

Sakr, Sherif, and Amal Elgammal. 2016. “Towards a Comprehensive Data Analytics Framework for Smart Healthcare Services.” *Big Data Research* 4: 44–58. http://dx.doi.org/10.1016/j.bdr.2016.05.002.

Saldivar, Alfredo Alan Flores et al. 2016. “Identifying Smart Design Attributes for Industry 4.0 Customization Using a Clustering Genetic Algorithm.” *2016 22nd International Conference on Automation and Computing, ICAC 2016: Tackling the New Challenges in Automation and Computing* (September): 408–14.

Sethi, Pallavi, and Smruti R. Sarangi. 2017. “Internet of Things: Architectures, Protocols, and Applications.” *Journal of Electrical and Computer Engineering* 2017.

Stimmel, Carol L. 2016. Big Data Analytics Strategies for the Smart Grid *Big Data Analytics Strategies for the Smart Grid*.

Stojmenovic, Ivan. 2014. “Machine-to-Machine Communications with in-Network

Data Aggregation, Processing, and Actuation for Large-Scale Cyber-Physical Systems." *IEEE Internet of Things Journal* 1(2): 122–28.

Sun, Jianing et al. 2015. "Smart Services for Enhancing Personal Competence in Industrie 4.0 Digital Factory." *Logforum* 14(1): 51–57.

Tang, Dunbing et al. 2018. "Using Autonomous Intelligence to Build a Smart Shop Floor." *International Journal of Advanced Manufacturing Technology* 94(5–8): 1597–1606.

Tao, Fei et al. 2014. "CCIoT-CMfg: Cloud Computing and Internet of Things-Based Cloud Manufacturing Service System." *IEEE Transactions on Industrial Informatics* 10(2): 1435–42.

Terkaj, W., and M. Urgo. 2015. "A Virtual Factory Data Model as a Support Tool for the Simulation of Manufacturing Systems." *Procedia CIRP* 28: 137–42. http://dx.doi.org/10.1016/j.procir.2015.04.023.

Wang, Jinjiang et al. 2018. "Deep Learning for Smart Manufacturing: Methods and Applications." *Journal of Manufacturing Systems* 48: 144–56. https://doi.org/10.1016/j.jmsy.2018.01.003.

Wang, Jinjiang, Peilun Fu, and Robert X. Gao. 2019. "Machine Vision Intelligence for Product Defect Inspection Based on Deep Learning and Hough Transform." *Journal of Manufacturing Systems* 51(March): 52–60. https://doi.org/10.1016/j.jmsy.2019.03.002.

Wang, Shiyong et al. 2016. "Towards Smart Factory for Industry 4.0: A Self-Organized Multi-Agent System with Big Data Based Feedback and Coordination." *Computer Networks* 101: 158–68.

Wang, Yichuan, Lee Ann Kung, and Terry Anthony Byrd. 2018. "Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations." *Technological Forecasting and Social Change* 126: 3–13. https://doi.org/10.1016/j.techfore.2015.12.019.

Wohlin, Claes. 2014. "Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering." *ACM International Conference Proceeding Series*.

Yan, Ye, Yi Qian, and Hamid Sharif. 2011. "A Secure Data Aggregation and Dispatch Scheme for Home Area Networks in Smart Grid." *GLOBECOM - IEEE Global Telecommunications Conference*: 1–6.

Yang, Dong, Ming Dong, and Rui Miao. 2008. "Development of a Product Configuration System with an Ontology-Based Approach." *CAD Computer Aided Design* 40(8): 863–78.

Yu, Cui, and Josef Boyd. 2016. "FB+-Tree for Big Data Management." *Big Data Research* 4: 25–36. http://dx.doi.org/10.1016/j.bdr.2015.11.003.

Zhang, Yan et al. 2012. "Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid." *IEEE Network* 26(3): 6–13.

Zhang, Yingfeng et al. 2018. "Coloured Petri Net-Based Active Sensing System of

Real-Time and Multi-Source Manufacturing Information for Smart Factory." *International Journal of Advanced Manufacturing Technology* 94(9–12): 3427–39.

Zhong, Ray Y. et al. 2015. "A Big Data Approach for Logistics Trajectory Discovery from RFID-Enabled Production Data." *International Journal of Production Economics* 165: 260–72.