An Argumentation Reasoning Approach for Data Processing

Erisa Karafili*, Konstantina Spanaki†, and Emil C. Lupu*

*Department of Computing, Imperial College London,
180 Queen's Gate, SW7 2AZ, London, UK
† School of Business and Economics, Loughborough University,
LE11 3TU, Leicestershire, UK
e.karafili@imperial.ac.uk, k.spanaki@lboro.ac.uk,
e.c.lupu@imperial.ac.uk

Abstract. Data-intensive environments enable us to capture information and knowledge about the physical surroundings, to optimise our resources, enjoy personalised services and gain unprecedented insights into our lives. However, to obtain these endeavours extracted from the data, this data should be generated, collected and the insight should be exploited. Following an argumentation reasoning approach for data processing and building on the theoretical background of data management, we highlight the importance of data sharing agreements (DSAs) and quality attributes for the proposed data processing mechanism. The proposed approach is taking into account the DSAs and usage policies as well as the quality attributes of the data, which were previously neglected compared to existing methods in the data processing and management field. Previous research provided techniques towards this direction; however, a more intensive research approach for processing techniques should be introduced for the future to enhance the value creation from the data and new strategies should be formed around this data generated daily from various devices and sources.

Keywords: Data processing, data quality, usage control, argumentation reasoning, data manufacturing, case scenarios.

1 Introduction

Big Data paradigm as a shifting phenomenon [14,33,7,40] provides access to a large pool of data resources; coupled with new storage, management and analytical techniques. These unprecedented opportunities arising from 'Big Data' are stressing firms to position themselves within this highly competitive data-intensive scheme, altering the way they generate, collect and transform data to actionable knowledge [10]. Gaining a competitive posture requires more than analytical techniques [36]; making sense [69] of this data is the challenge of organisations, forming a new way of data-based decision-making, disrupting the business landscape while moving from the world of "making things" to a "world

of outcomes" [44]. The industrial world was disruptively changed during the last century; while the manufacturing process has gone through multiple revolutions (agricultural, industrial, digital etc.) and respectively multiple alterations in the way products and services are formed. A critical view of the literature associated with the manufacturing context [41,62] of the last decades, reveals that this evolution around data has influenced radically also the production, distribution and supply chain world.

The data evolution (also often referred as "Big Data" evolution) has also accelerated the popularity of service industries around data (for storing, analysing, processing etc.). In the context of these data services; myriads of data are shared as well as stored, used and transformed, not only by users but also by companies, organisations, and governments. Security concerns regarding shared data, e.g., privacy, confidentiality, integrity, and the necessity of protecting them are becoming serious issues. Towards this direction, the regulatory and legislative environment is continuously updated for meeting citizens' rights, protecting their data privacy, anonymity, and security. The regulatory and legislative implications of security issues are becoming rather more severe when we extend the scope across the immediate organisational, governmental or even country borders; where multiple legislative domains are applied, and frequently they lead to conflicting behaviours and interests. Exchanging data implies that all parties agree on the associated rules to be enforced; this is often referred as a data sharing agreement [59] (DSA). The DSA can be seen as a contract, between two or more parties, and the different rules are the terms of the contract. The terms express how and who is permitted or denied to access, delete, use, and share the data, along with the different constraints that should be respected. The enforcement is done when the data are used, and it evaluates the requests and usage of the data against the set agreements.

Constructing and representing DSAs is not trivial, as they should incorporate data access and usage rules for the security and privacy of the data, as well as users preferences, business and legislative rules applicable to that case. All the above rules are applied to the same bunch of data; therefore there can be conflicting behaviours between rules, and differentiated legislative, regulatory rules and domain contexts that bring inconsistencies inside the DSAs rules. Different techniques [42,63,71] were introduced for data services towards various security solutions, with a major focus on data access permissions and an efficient usage control [17,70], suitable for stored or integrated data from multiple parties and sources. Access and usage control is a well-studied research area [39,53], however the existing solutions do not permit a fine grained representation of the different types of rules and their constraints, as well as the associated conflict detection and resolution.

The aim of this study is to propose a novel data processing technique using a policy analysis language for the representation of the sharing agreements and quality attributes of the data. We extend our research scope in an industrial context where the data quality control should also be studied as it is often a neglected aspect in the context of data sharing agreements. Initially, we provide

the theoretical background where this study was based, coupling two streams of literature. Previous research in data management was examined in order to explain the data manufacturing analogy and the quality problem which was also highlighted in past decades, as well as the data access and usage control background for building the concepts associated with DSAs. The following sections develop the proposed methodological approach based on a policy language for argumentation and abductive reasoning used in the data processing context, and its explanation through a use case scenario and the relevant analysis and representation. Our study ends up with a research agenda for the future implications of this approach, as well as how this can be extended and applied in industrial data-intensive environments.

2 Theoretical Background

The analogy of the data processing to the manufacturing processing of physical materials is prevalent in the literature of data management. There are many similarities between the mechanisms of data processing and the manufacturing processing; however, there are some significant differences. In a manufacturing process, physical materials are input into a process, the materials are transformed, and the resulting output is a manufacturing product. In the data processing, the data represent the input into the processing mechanism, and a transformed data product is the output of the process. Data of bad quality used through the data process will remain bad until the quality is improved (until the problem is actively cleaned up or removed). Data sharing controls were not well-studied in previous decades, as the data were mostly shared within the boundaries of a company or between single databases, where the trust and security issues were solved by the individual sharing entities and the associated agreements between interested parties.

The data manufacturing framing presented in data management literature of the previous decades should be extended in a context across boundaries between individuals, firms and countries and the target should be tailored data products, as these are mostly the results of the data evolution. Data manufacturing analogy was based on the data artefact as a unit of analysis; however, data era requires novel techniques focusing on the processing of data but not solely on their processing mechanism, but also the quality and sharing attributes associated with them. This section will explore the previous research of data management and data sharing literature, with the view to provide a background for the problem area and how this is expanded in a data-intensive industrial context.

2.1 Data Management and Quality Attributes

Plenty of examples nowadays can provide evidence that companies from multiple industries invest in data management solutions, with the view to improve and expand the production of enterprises through the use of their analytical skills, or by viewing and optimising their supply chains of their core business [62,61].

Although data can be used along with the core business focus in different industries, the recent data evolution expanded the business scope and disrupted the operating models providing opportunities to process this data, create new data and information products/services and also to resell and exchange this data. Reviews of the literature in the context of data management in an industrial context mostly focus on "Supply Chain Analytics (SCA)" [64] as a way in "developing supply chain strategies and efficiently managing supply chain operations at tactical and operational levels" [64]. Supply Chain Management (SCM) focuses mostly on how the analytics can be applied to strategic decisions related to SCM [64,62], how efficiency and effectiveness of supply chains can be improved through the use of data [61] as well as the data strategies and servitization around supply chains [46]. This direction reflects that the research focuses mostly the use of data within an industrial context; nevertheless, our research focus will be on data processing techniques, setting data manufacturing analogy from data management literature as the main theoretical background, analogous to the product/service manufacturing processes. Literature reviews and frameworks referring to data/information processing are presented and summarised in Table 1, in Appendix A. This summary reveals that data analogous to physical materials are moved through a manufacturing process which reshapes/reconfigures them in information/data products.

Data processing was initially introduced by Brodie [6] through the analogy between product manufacturing and data manufacturing when data quality was a primary concern in transforming data to valid information and knowledge [1]. Some of the most indicative studies around these areas developed the concept of data manufacturing analogy in order to find out the path for better data quality [38,1,18] and they provided frameworks to describe and track data manufacturing process [65,2,66,56]. A simple framework of input-process-output describing the similarities between the two manufacturing processes was proposed in [66] and calls for continuously defining, measuring, analysing, and improving data quality. Mostly, the data manufacturing analogy was focusing on data quality and the ways to ensure that we can trust the data we use in manufacturing processes. Recent studies following the data manufacturing path can be considered those in [19] and [16] where they introduce the need for continuous improvement in the SCM data production process, suggesting a framework for establishing a data quality control mechanism. These two studies are investigating the data manufacturing process via a data quality lens and expand the research focus to new topics related to the processing of data, and how to improve the quality of this data with the use of various techniques and methods.

2.2 Data Access and Usage Control

Research around data sharing has provided data-centric solutions for protecting the used and shared data, aside from focusing on protecting the databases where they are stored [15], the network used for their transfer [28], or constructing coordination techniques for data re-use [25]. The focus in data-centric solution is due to the increase in the connectivity and the diversity of attacks.

Protecting and ensuring the security of all the environments where the data are collected, processed, and shared is a major challenge for all the interested parties. Therefore, data-centric security solutions have an important position in the literature [4,31,42,63,71] with focus on protecting data transfers and transactions. Data-centric security solutions present two main challenges: the control of data access and the usage of data. Both of them, have been widely studied and various solutions are developed for solving these problems [34,12].

Previous approaches in role-based access control [12] are based on different user roles for data access controls. Motivated by this direction, we will expand these aspects with a data representation technique for different user roles with specific access and usage policies for the data. Usage control (UCON) [48] is a widely studied concept following different approaches, for controlling the access and usage of digital information emphasising the problem of rights delegation [47], or a two level policy language that represents the notions of prohibition and obligation, and a generic server-side architecture for UCON [52]. Applications of UCON are also presented in the context of distributed systems [29], when having a data flow in-between different connected systems, or in the context of multiple distributed systems as a fully decentralised infrastructure for enforcement of global usage control [30]. Our work is building on the previous research of UCON while expanding the scope regarding the policies that represent permissions, denials, obligations and delegations of rights.

Another interesting approach to sharing and accessing data is the use of sticky policies [42,43]. Sticky policies are machine readable policies that contain conditions and constraints attached to data that describe how the data should be treated while shared among multiple parties. The sticky policy paradigm [27] and technologies for enterprise privacy enforcement and exchange of customer data are represented through a privacy control language [26] for specified privacy rights and obligations. The privacy control language presents authorisation management and access control for user consents, obligations and distributed administration, with the extension of the sticky policy paradigm also for the cloud environment [49,60]. Our research will build on the policy language for policy representation and will expand the application of this technique in the industrial use of data.

Data usage concerns are usually expressed by the different entities using the data. These entities, before creating, sharing and using the data, should agree regarding the different rules that describe how the data should be treated, called data sharing agreements [59] (DSAs). The DSAs describe not only the agreements between the involved parties but also the compliance to the different business, legislative and regulatory rules for the various contexts of the data sharing. A language representation of different rules for data sharing agreements (DSAs) as presented by [39] fails to provide expressivity. This language cannot permit the representation of complex DSAs, as well as analysis for the DSAs and leaves unsolved the problem of deciding which rules to apply to the DSAs.

All the above-represented approaches, from the data access and usage control to the sticky policies and finally the DSAs representation, seem incomplete to

provide a decision background for the rules that should apply to the shared data. Following this motivation, we propose a combinatory analysis of the rules with a conflict resolution technique. The proposed analysis is an enhancement of the one introduced in [24,23], as we enrich it with data quality analysis that can be easily extended for big data applications. Our solution is based on abductive [22] and argumentation based reasoning approach [5,11], as this technique can facilitate decision making mechanisms [21,3] under conflicting knowledge.

2.3 Introduction to the Use Case Scenario

In this work, we show our data processing technique based on DSAs in a realistic use case. Let's introduce our scenario, taken from an e-health example of a European Project (Coco Cloud project[1]). In this use case, the data needs to be collected, processed and shared between different actors, e.g., data subject, data controller, recipients, and data processor. The actors need to stipulate agreements between each other that describe how the data should be treated, the data sharing agreements, composed of security, legislative and business rules. Deciding the rules to apply to particular cases is not trivial, as various conflicts might arise. The conflicts need to be captured and solved.

One of the main actors is the *data subject*, in our use case is the patient, some of his rights are to access/delete his medical data and to know who is processing his data. The *data controller* is an entity (public authority, agency, legal person), which determines the purpose and means of processing the data of the data subject. In our use case, the hospital is the data controller of the patient's data and determines the purpose for which the data are processed (e.g., administrative purpose or treatment purpose). The doctors of the hospital are the *data recipients* that need to comply with the data controller rules. The data recipients are considered as part of the data controller, the employees within the hospital do not stand as separate entities than the hospital itself. The hospital that is the data controller has various rules of how the doctors can access the patient's data, e.g., a doctor needs to be inside the hospital for accessing the patient's data (geographical constraint), he needs to be during his office hours (temporal constraint), and he needs to be the patient's treating doctor (role-based constraint).

The *data processor* is an entity (public authority, agency, legal person) that is processing the data on behalf of the controller. In our use case, the cloud provider is considered the data processor as far as it respects the instructions of the controller. The controller rules that should be respected by the processor can also have a legal nature, e.g., if the controller is in an EU country, the cloud provider should as well be in an EU country and cannot send the data to countries outside the EU and EEA.

A *third party* is an entity (public authority, agency, legal person) that is not the data subject, data controller or processor, and that under the direct authority of the controller or processor is authorised to process the data. In our

[1] `http://www.coco-cloud.eu/`

use case, a doctor outside the controller hospital is considered a third party. Once access is obtained, the third party becomes a data controller and has to comply with the data protection principals. The patient can be in different situations, e.g., intensive treatment, unconscious, emergency, that affect how the data are shared and used.

3 Methodology

The study presents the proposed approach, where the bunches of data are processed from different entities. The latter establish different agreements between each other, called data sharing agreements. The DSAs are composed of various constraints and rules. We use an expressive policy analysis language for representing the DSAs. An important aspect of the data processing mechanism is data quality. We enrich the policy language to capture various data quality properties like accessibility, timeliness and accuracy. The used policy language permits the analysis of the various policies and the detection of the rising conflicts, redundancies or the missing cases. The proposed method implements an analysis and representation based on argumentation and abductive reasoning to capture and solve conflicts between context dependent rules. The introduced analysis permits the construction of correct and efficient DSAs that apply in different contexts during the processing of data.

3.1 Policy Language for DSA and Data Quality Representation

The proposed model is based on the policy analysis language [8] that is constructed using the Event Calculus [32]. This language represents the required rules and constraints during the access, usage and sharing of data. The policy regulation rules are composed of predicates and domain description ones, and represent authorisation and obligation rules, and have in their structure *subject*, *targets*, and *actions*[2]. Some of the predicates of the policy language are introduced below.

$$req(Sub, Tar, Act, T) \qquad obl(Sub, Tar, Act, T_s, T_e, T)$$
$$permitted(Sub, Tar, Act, T) \qquad denied(Sub, Tar, Act, T)$$

The above predicates represent correspondingly: a request made from the subject Sub, at the instant of time T, to perform a certain action Act at the target Tar; the obligation for a given subject to perform an action during the period of time from T_s to T_e; that a given subject is permitted/denied at time T, to perform a certain action to the target. A domain description predicate is *holdsAt*, which means that a given property/predicate is true at a given instant of time. The used policy language can represent the permission, denial and obligation concepts for the DSAs. In the following examples, we introduce the representation of some DSAs rules from our use case.

[2] For the sake of space, we give a brief introduction to the language. For a detailed representation of the language, we address the reader to [8].

Example 1. Bob (B) is the family doctor ($fDoc$) of the patient Alice. The family doctor has the permission to access Alice's prescriptions[3], (A_presc), at the instant of time T.

$$permitted(B, A\_presc, access, T) \leftarrow holdsAt(fDoc(B, Alice), T), \\ holdsAt(owner(Alice, A\_presc), T).$$

Example 2. The hospital (H), where Alice was hospitalized ($hosp$), needs to delete her data at maximum 2 years (700 days) after she was discharged, ($disc$), from the hospital.

$$obl(H, A\_presc, delete, T_2, T, T_1) \leftarrow holdsAt(hosp(Alice, H), T_1), \\ holdsAt(disc(Alice, H), T_2), \\ holdsAt(owner(Alice, A\_presc), T_2), \\ T_1 < T_2,\ T > T_2,\ T \leq T_2 + 700.$$

Example 3. Bob can access Alice's prescriptions when he is inside the hospital. For ensuring that Bob is inside the hospital we use $pos(Bob, Location)$ that gives Bob's location, and $hospP(Hospital, Location)$ that gives the hospital geographical location. Thus, we use the function $same(L_1, L_2)$ that checks if the two given locations L_1, L_2, are the same or are geographically nearby each other, to be considered the same location.

$$permitted(B, A\_presc, access, T) \leftarrow holdsAt(fDoc(B, Alice), T), \\ holdsAt(owner(Alice, A\_presc), T), \\ holdsAt(work(B, H), T), \\ holdsAt(hospP(H, L_1), T), \\ holdsAt(pos(B, L_2), T), same(L_1, L_2).$$

Ensuring Data Quality The used policy language permits the representation of different predicates, authorisation, and obligation policies, together with domain description policies. The sharing and usage of data bring the need of describing other properties related to the quality of the collected data. When working with data quality the entities that are using, sharing, storing the data are called *data consumers*. In our case, the data consumers are considered the data collectors and the data processors. The data quality is a major factor when we are working with data consumers, where *data quality* is defined as data that are fit to be used by data consumers [67]. Through our DSAs representation and the applied policies, we grant an important characteristic of data quality, which is the data *accessibility*. We ensure the data accessibility that complies with the constraints for accessing the data.

Timeliness/Freshness Another important data quality characteristic is data *timeliness*, or better the degree to which data represent reality from the required point in time when the event occurred. For the context we are working on, generally, the data have a specific subject and a data controller that together

[3] We use the *owner* property for relating the patient to his data.

with the various legislations and business rules can decide the different policies to be used for accessing and using the data. For representing the data timeliness, we use the concept of data *freshness*. Data freshness is a predicate that expresses the last time when a given piece of data has been updated: $freshness(Tar, T)$, where Tar is the targeted data, and T is the last instant of time when the data were updated. Through the use of the $freshness$ predicate, we are able to represent the data timeliness depending on the different contexts. Let's see how we can apply the above predicate to our use case.

Example 4. Suppose that a patient is hospitalized for a particular illness or symptoms related to an illness that is still being studied. He is offered to be part of a pilot project, and give the consent of sharing his medical data (anonymized/ sanitised) with a team of researchers. Therefore, the patient will be monitored/ visited in specific instants/intervals of time, e.g., the doctor should visit, ($visit$), the patient every four hours. For granting the data timeliness[4], every time the patient is visited/monitored his medical data are updated, upd. In case, there are no changes to the patient's data (e.g., same vital functions) the old data are confirmed for that instant of time.

$$\begin{aligned} freshness(Tar, T) \leftarrow\ & fulfilled(Sub_1, Tar, upd, T', T_e, T'), \\ & holdsAt(visit(Sub_1, P), T'), \\ & holdsAt(owner(P, Tar), T'), \\ & \textbf{not}\ holdsAt(visit(Sub_2, P), T''), \\ & freshness(Tar, T'''),\ T_e > T > T',\ T' > T'' > T'''. \end{aligned}$$

The above predicate states that the given data (Tar) at the instant of time T are fresh, as the data of the patient are updated, every time the patient is visited. A deontic obligation holds for the doctor to update the patient's medical data every time he visits the patient.

$$\begin{aligned} fulfilled(Sub, Tar, T, T_e, T) \leftarrow\ & obl(Sub, Tar, upd, T, T_e, T), \\ & holdsAt(Sub, Tar, upd, T). \end{aligned}$$

$$\begin{aligned} obl(Sub, Tar, update, T, T_e, T) \leftarrow\ & holdsAt(Sub, P, visit, T), \\ & holdsAt(owner(P, Tar), T), T_e > T. \end{aligned}$$

The above predicates ensure that every time a patient is visited, the doctor updates the patient's data. Thus, the data freshness is in line with the visiting time, and the data timeliness is respected.

The timeliness of the patient's data described above is linked to the actions of actors like doctors and nurses. The patient can also be monitored by medical IoT devices that measure, ($meas$), different vital functions. In this case, for ensuring the timeliness of the data, we expect that the patient's medical data are collected

[4] We use the predicate $fulfilled(Sub, Tar, Act, T_s, T_e, T)$ that denotes that an obligation for a subject to perform an action to the target, from T_s to T_e, is fulfilled.

and saved simultaneously. The latter can be ensured by the below predicates[5].

$$\begin{aligned} freshness(Tar, T) \leftarrow\ & fulfilled(Sub, Tar, upd, T, T, T), \\ & holdsAt(meas(Sub, P), T), \\ & holdsAt(owner(P, Tar), T), \\ & \textbf{not}\ holdsAt(meas(Sub, P), T''), \\ & freshness(Tar, T'''),\ T > T'' > T'''. \end{aligned}$$

$$\begin{aligned} fulfilled(Sub, Tar, T, T, T) \leftarrow\ & obl(Sub, Tar, upd, T, T, T), \\ & holdsAt(Sub, Tar, meas, T). \end{aligned}$$

$$\begin{aligned} obl(Sub, Tar, upd, T, T, T) \leftarrow\ & holdsAt(Sub, P, meas, T), \\ & holdsAt(owner(P, Tar), T). \end{aligned}$$

Linked Data An important problem related to the data sharing is the linked data problem. One of the main facets of this problem is deciding the policies that should be applied to the new data, that were produced by processing existing ones. The new data are linked to the existing ones, and deciding their policies is not trivial, as processed data can have different policies, also in the conflict with each other, and the nature of new data is not known. The solution we propose is the use of the *type of data*. While working on our use case, the data are divided into different types, e.g., personal information, prescriptions, private prescriptions, and every type has its corresponding sets of policies. When new data are created, after the processing of old ones, the data subject and recipients (e.g., patient, doctor) are the ones that decide the type of data and consequently their corresponding set of policies. In this particular case, the responsibility of deciding the policies for the new data is given to the data subject and recipient that often can suffer from human errors. Despite the human errors, we believe this would be a good solution, especially for the e-health scenario, where a division of access related to the security type of data cannot be applied, e.g., the doctor can process private prescriptions of the patient (e.g., depression treatment) for producing routine prescriptions of the patient (prescription with a low privacy level).

Another interesting future solution, for the linked data problem, would be the extension of an *audit process* [50] which verifies the use of data for the intended purpose. Different tags should be added to the audit system and the purposes. Thus, if the audit process confirms the purpose of the data, the policies associated with that purpose are applied to the data.

Data Accuracy Another important property of data quality is the *accuracy* of the collected data. When the data collection is made by a human actor, we can put in the act a deontic obligation for the actor to use a particular accuracy when getting and saving the data. This case can suffer from human errors. On

[5] In the previous case, when the doctor visits the patient, we give an interval of time $T - T_e$ to him, for updating the data. We expect IoT devices to make a live update of the data, without the need of extra time.

the other hand, when a device is collecting the data, e.g., an IoT device, we can be more specific and ensure the data accuracy by checking the parameters of the various devices.

$$accuracy(Tar, T) \leftarrow holdsAt(meas(Sub, P), T), \\ holdsAt(acceptP(Sub), T), \\ holdsAt(owner(P, Tar), T).$$

In the above case, we state that the data collected for a particular subject are accurate, as the device that collected them is able to measure the data with acceptable parameters, ($acceptP$), for ensuring data accuracy. The same can be stated for a doctor visiting the patient and updating the data with an acceptable accuracy.

3.2 Data Processing with Argumentation Reasoning

The data are collected, used and shared between different entities. The rules to be applied while processing the data are given through a decision process, which given the data, the various entities together with their preferences, and the applicable legal, security and business rules, decide the rules that apply to particular contexts. As introduced above these rules are called data sharing agreements, and are represented as policies. Due to the heterogeneity of the rules, and their context dependability the policies that represent them can be in conflicts, redundant, or not complete. We introduce a policy analysis for capturing the above, solving the various conflicts, and perform a decision process. The introduced analysis permits to implement and accomplish an efficient data processing.

The policy language enables an analysis based on an abductive constraint logic programming system, A-system[6] [45], which is performed to the rules and permits their efficiency and soundness. In particular, the *modality conflicts* analysis task finds conflicts between policies regulation rules, and permits to have sound DSAs. It can capture the case when an action is both permitted/obliged and denied on the same instant of time, as well as more complex conflicts. The *coverage of gaps* analysis finds the different gaps (cases) that are not covered by the DSAs rules, and permits the construction of a complete list of rules that should be part of the DSAs, e.g., when there is an explicit request from a subject to perform an action, and there is no authorization policy rule that neither permits nor denies this request. The *policies comparison* analysis checks whether a policy is included/equivalent/implied by another one. This analysis improves the efficiency of the DSAs, by identifying redundant rules, that can be easily removed from the DSAs.

The above analysis cannot capture the *conceptual conflicts*, as the latter are not direct conflicts between predicates (e.g., permitted/obliged and denied predicates) and are context dependent. The rules can be in conflict with each other, as they might hold for general domain description predicates but not for specific ones, or vice versa. The resolution of the conceptual conflicts is done

[6] A-system `http://dtai.cs.kuleuven.be/krr/Asystem/`

by introducing priorities between rules [3], and is a decision-making problem. Techniques based on argumentation reasoning [21,5], (that is a non-monotonic reasoning [11]) together with abductive one, is introduced to find and solve the conceptual conflicts. Argumentation reasoning implements decision-making mechanisms for conflicting rules that have priorities/preferences between them and that is strongly context dependent. Argumentation reasoning permits to represent the various conflicting rules, the context where they are valid and the preferences between them. The priorities between rules permit us to work with conflicting policies and to analyse them. The above analysis is implemented using GorgiasB[7] [57], which is a tool for preference-based argumentation with a graphical user interface.

Our decision-making technique has as input the rules together with the domain description predicates that can be facts or defeasible knowledge, and finds the conflicts between rules, if there is any, and solves them. The resolution of the conflicts is made by introducing priorities between rules, called *priority rules*, and explicitly specifying when a rule has to be considered stronger than another one. A preference/priority relation, denoted by $>$, is used to indicate preferences between rules. Given two conflicting rules r_1 and r_2, where for the context and the information we have, r_1 should be applied instead of r_2, we denote it with $r_1 > r_2$. The introduced priority rules together with the existing rules are checked, and if any conflict is found, other priorities rules are introduced.

For every context and information given, the above analysis and conflict resolution provide the DSAs that apply to the data. The DSAs are used as regulatory rules between entities while processing the data and permit the compliance of all the various requirements and constraints: security, business, and legal.

4 Case Representation and Analysis

In this section, we continue with our use case that was already introduced in the previous sections. In our use case, we deal with an EU e-health scenario, where we want to model the data sharing agreements between different entities, for sharing and using the data. The different entities are the patients $\mathcal{P} = \{P_1, P_2, \cdots\}$, the service providers that are the hospitals $\mathcal{H} = \{H_1, H_2, \cdots\}$, and the doctors $\mathcal{D} = \{D_1, D_2, \cdots\}$, that work in hospitals. Every patient has his associated *data*. The data can be of three types:

- prescriptions: $Presc(data)$, e.g., blood pressure, analyses, x-rays;
- private prescriptions: $PData(data)$, e.g., anti-depressive treatments;
- personal information: $PInfo(data)$, e.g., contact and emergency contacts.

Every patient (P) has a *family doctor*: $FDoc(D, P)$. When the patient is treated/hospitalized in an hospital he has also the *treating doctors*: $TDoc(D, P)$. In this case, for D to be the treating doctor of P, then D should work in the same hospital where P is treated/hospitalized, as follows:

$$TDoc(D, P) \text{ when } Hosp(P, H) \wedge Work(D, H).$$

[7] GorgiasB `http://gorgiasb.tuc.gr/`

The patient's data are used, accessed, and shared between different entities by respecting their DSAs. The first step is to agree on the terms of the DSAs, where some DSAs terms, usually legal ones, are irrefutable. Let us give some of the DSAs terms for our scenario.

1. The family doctor can access to all the data of the patient.
2. The patient can access to all his data.
3. The treating doctor can access to the prescription data of the patient.
4. The hospital regulation says that the treating doctor can access the patient's data during his working time, and while he is in the hospital.
5. The treating doctor cannot access the other types of the patient's data (private prescriptions and personal information), and cannot access to all types of data when he is not in the hospital, or not during his shift.
6. Nobody else can access the data.

The family doctor accesses to all the patient's data, described as follows[8]:

$$Access(data, P, permitted) \leftarrow FDoc(D, P) \wedge Owner(P, data). \tag{1}$$

Rule 2 states that the patient can access to all of his data:

$$Access(data, P, permitted) \leftarrow Owner(P, data). \tag{2}$$

Rule 3 states that the treating doctor is permitted to access the patient's prescriptions. Rule 4 states that the treating doctor can access the patient's data during his shift, $shift(Doctor)$, and while he is in the hospital. For ensuring the latter, we use $hospP(Hospital, Location)$ and $pos(D, Location)$. The other accesses, e.g., when the doctor is not in the hospital, not during his working hours are not allowed, rule 5. Below, we represent rule 3 and 4 together, and rule 5.

$$\begin{aligned} Access(data, D, permitted) \leftarrow\ & TDoc(D, P) \wedge Owner(P, data) \wedge \\ & Presc(data) \wedge shift(D) \ \wedge pos(D, L_1) \wedge \\ & hospP(H, L_2) \wedge \ same(L_1, L_2) \end{aligned} \tag{3}$$

$$\begin{aligned} Access(data, D, denied) \leftarrow\ & TDoc(D, P) \wedge Owner(P, data) \ \wedge (PData(data) \\ & \vee \ PInfo(data) \vee \textbf{not} \ shift(D) \vee (pos(D, L_1) \wedge \\ & hospP(H, L_2) \wedge \textbf{not} \ same(L_1, L_2))) \end{aligned} \tag{5}$$

A patient might be in an emergency situation (means a high level of risk for his life): $Emerg(P, H)$. When the patient is an emergency situation, the policies for the legislative and business rules are as below.

7. The treating doctor can access the prescriptions of the patient, when the patient is in an emergency situation, and the doctor is in the hospital and during his shift.
8. The treating doctor can access the personal information of the patient, e.g., for notifying the family members, when the patient is in an emergency situation, and the doctor is in the hospital and during his shift.

[8] We represent the rules with the use of a semi-natural language.

9. The treating doctor cannot access the patient's private prescription, when the patient is in an emergency situation.

$$\begin{aligned} Access(data, D, permitted) \leftarrow\ & Emerg(P, H) \wedge TDoc(D, P) \wedge \\ & Presc(data) \wedge Owner(P, data) \wedge \\ & pos(D, L_1) \wedge hospP(H, L_2) \\ & \wedge\ same(L_1, L_2) \wedge shift(D) \end{aligned} \tag{7}$$

$$\begin{aligned} Access(data, D, permitted) \leftarrow\ & Emerg(P, H) \wedge TDoc(D, P) \wedge PInfo(data) \\ & \wedge\ Owner(P, data) \wedge pos(D, L_1) \wedge hospP(H, L_2) \\ & \wedge\ same(L_1, L_2) \wedge shift(D) \end{aligned} \tag{8}$$

$$\begin{aligned} Access(data, D, denied) \leftarrow\ & Emerg(P, H) \wedge TDoc(D, P) \wedge \\ & PData(data) \wedge Owner(P, data) \end{aligned} \tag{9}$$

Our analysis finds that rule 7 is a sub-case of rule 3. Thus, we can remove rule 7. Rules 8 is in conflict with rule 5. This type of conflict is captured by the argumentation reasoning analysis. For solving this conflict, we put a preference relation stating that the emergency situation has higher priority. Thus, the doctor can access the personal information of the patient, represented as rule 8 $>$ rule 5. The redundancy analysis finds that rule 9 represents the same policy as rule 5, where actually the latter is more generic. For improving the efficiency of our DSAs, rule 9 can be removed.

As described in the previous sections, the patient (P) may decide to be part of a research study for a particular disease, $Study(P)$. In this case, his data are shared with third parties, e.g., research institutes, insurance company, government entities. Some of the uses that these entities can do with the data are for studies in particular diseases, creating adequate health campaign or stipulating new insurance plans. How and to whom the data are shared depends on their freshness, ($Fresh$), and accuracy, ($Accurate$), and the entity to whom they are shared. The entities are divided into *basic* (e.g., insurance company), *silver* (e.g., government entities), and *gold* (e.g., research institutes) members, depending on the type of agreements they have with the data subject and controller. The rules that apply in the patient data are as below[9]:

10. The gold members ($Gold$) can access to the patients normal and private prescriptions that are fresh and accurate.
11. The silver members ($Silver$) can access to the patients normal and private prescriptions that can be accurate or fresh.
12. The basic members ($Basic$) can access to the patients normal prescriptions that are not accurate, but can be fresh.

[9] For sake of simplicity, we omit the data sanitisation process, but we assume that all the shared data during the study are sanitised.

For the last case, if the data is accurate, an impoverishment process takes place, that takes part of the data accuracy out, ($Cast$).

$$Access(data, M, permitted) \leftarrow Gold(M) \wedge Owner(P, data) \wedge Study(P) \wedge (Presc(data) \vee PData(data)) \wedge Accurate(data) \wedge Fresh(data) \quad (10)$$

$$Access(data, M, permitted) \leftarrow Silver(M) \wedge Owner(P, data) \wedge Study(P) \wedge (Presc(data) \vee PData(data)) \quad (11)$$

$$Access(data, M, permitted) \leftarrow Basic(M) \wedge Owner(P, data) \wedge Study(P) \wedge (\textbf{not}\ Accurate(data) \vee Cast(data)) \wedge Presc(data) \quad (12)$$

The above rules are in conflict with rule 6, where nobody else except the doctors can access the data. As the patient is part of the study, the above rules are stronger then rule 6, represented as rule 10 > rule 6, rule 11 > rule 6, and rule 12 > rule 6.

The legislation says that the medical data of the patient can be shared just between entities inside the EU or EEA. Thus for the above case we have that[10]:

13. If a member that is part of the study is not part of the EU/EEA, than it cannot access the patient's data.

$$Access(data, M, denied) \leftarrow (Basic(M) \vee Silver(M) \vee Gold(M)) \wedge Owner(P, data) \wedge Study(P) \wedge \textbf{not}\ EU^*(M) \quad (13)$$

This rule is a sub-case of rule 6, but is stronger then the above rules, as is a legal requirement that should be respected. Thus, rule 13 > rule 10, rule 13 > rule 11, and rule 13 > rule 12.

5 Conclusion and Future Research Directions

The purpose of this research was to extend the 'data manufacturing' concept of previous decades, to a data-intensive environment across organisational, individual and country boundaries, where 'data products' are accessible to different entities who have the approved rights on them. By using the data processing approach, we unfold the potentialities of data marketplaces, where tailored "data products" can be created, shared, sold, used or even accessed by various entities. We also identify the emerging areas of interest arising within the context of "Big Data". Within this context, we argue that data can be processed and can create value through tailoring techniques across organisational boundaries with the help of DSAs and usage control rules while developing data products/services as well as disrupting existing business models to facilitate such a change.

[10] A predicate $EU^*(M)$ is used to check if the given entity is located in an EU/EAA country or not.

Future research should consider that data nowadays are multiform and multi-source; therefore new approaches are required for value extraction appropriate for any of these forms, creative enough for innovative industrial usage and extensive enough to process the masses of heterogeneous data. Data processing and manufacturing approaches should be investigated further for the path to better data quality, along with new frameworks to describe and track data processing in different industrial applications. Except for data quality, data privacy is an emerging concern, as often serious threats arise when the datasets are shared among third parties. Thus, the ways to prevent such issues provided a new research agenda around trust and shared responsibility among the actors and entities involved. Furthermore, data collection, processing and storage techniques and methods is a progressively expanding research area, as there is a wide interest in how data can be generated and exploited and therefore what are the right tools and methods for that. Data generation and exploitation strategies can also focus on the organisational aspects as well as the capabilities and skills the firms should acquire for building innovation in data-intensive contexts. Moreover, a strategic way of coupling multi-source data in different innovative ways while creating data product/services and new value streams for the companies should be proposed and explored further the necessary capabilities, skills and innovative ways of handling and processing the data.

Acknowledgments

Supported by FP7 EU-funded project Coco Cloud grant no.: 610853, and EPSRC Project CIPART grant no. EP/L022729/1.

A Appendix

Table 1: Studies with focus on Data Manufacturing Process

Studies with focus on Data Manufacturing Process		
Study	**Description**	**Key points**
[66]	Analysis of data quality literature following "data manufacturing" analogy as a framework for literature synthesis	– Data quality issues – Data aspects – Data manufacturing/ processing – Data products
[1]	Review of the information manufacturing stages through a discussion, raising issues of data quality	– Extracting information from databases – Information manufacturing/processing – Data quality
[2]	Introducing the problem of "information produc" quality providing an assessment model of information quality (tracking information attributes-timelines, accuracy and cost)	– Information manufacturing/processing – Information products – Information manufacturing systems – Information quality – Data quality
[6]	Discussing the role of data quality for the effective use of information systems - examining the tools, concepts and techniques around data quality	– Data reliability – Data quality in programming and databases – Data quality assessment

[13]	Literature review on "data" definition (5 approaches) expanding the discussion to data quality issues, introducing 4 dimensions (accuracy, completeness, consistency, current-ness)	– Data definition – Data quality dimensions
[56]	Information as an inventory approach, involving a 3 stage process (raw materials-process-finished goods)	– Information processing/ manufacturing – Input-process-output for information manufacturing – Data as a key resource
[65]	Positioning the "data quality" problem in Total Data Quality Management (TDQM) perspective proposing a methodology around TDQM	– Information product – Information product characteristics – Information manufacturing – Information quality assessment – Information manufacturing systems
[66]	An attribute-based model for data quality assessment requirements analysis to serve quality indicator identification	– Data quality – Data manufacturing – Data management – Data quality requirements
[68]	A framework for data quality following a data consumer perspective - organizing data quality dimensions (intrinsic, contextual, representational, accessible). A survey approach is followed to test and refine the framework	– Data manufacturing systems – Data raw material – Data products – Data consumers
[58]	Conceptualizing data quality in a context of organizational processes, procedures, roles employed in collecting, processing, distributing and using data	– Data quality – Data from multiple sources – Databases – Information for decision making
[35]	Information quality assessment tool developed through surveys in 5 organizations	– Information quality – Information product – Information quality assessment
[51]	Data quality assessment metrics - highlighting the distinction between "data" and "information" (information is presented as processed data)	– Data and information – Processed data (information as an outcome) – Data quality dimensions
[20]	Information quality benchmarks for product and service performance introducing a measurement instrument for Information Quality dimensions	– Information quality benchmarks – Dimensions of Information Quality
[54]	Discussion about poor data quality and its impacts (operational, strategic, tactical) on the enterprises and their customers in the "Information Age"	– Information ecology – Data quality issues – Data accuracy
[9]	"Information Ecology" approach as a holistic view of the information environment (endogenous and exogenous), distinguishing "data", "information" and "knowledge" as distinct aspects	– Data storage – Databases mastering information complexity – Technology as enabler of information improvement – Data, information, knowledge
[55]	Methodologies for data quality programs in Information Age enterprises	– Aspects of data management – Management roles around data – Data quality design

[37]	Overview of data and information quality landscape providing a framework for categorization of topics and methods	– Data and information quality – Data quality methods – Data quality topics
[18]	Data quality attached to the development and delivery of products/services as a crucial quality factor - suggestions of a 4-part quality program	– Dimensions of data quality – Data tracking and quality control
[19]	Proposing the use of control charting methods for data quality monitoring - stressing the data quality problem as crucial in Data Analytics within 2010ies	– Data production process – Data quality – Data analytics
[38]	Decision-making based on integrated, high quality information (systems theory in decision-making) -highlighting themes for data warehousing decision-making (integration, implementation, intelligence, innovation)	– Data warehousing challenges – Input-process-output – Data-based decision-making – Aspects of data-based decision-making (integration, implementation, intelligence, innovation)
[16]	Ways of producing, organizing and analysing data - presenting data quality problem in the notion of SCM problems	– Data quality – Monitoring and controlling data quality

References

1. Stephen E. Arnold. Information manufacturing: the road to database quality. *Database*, 15(5):32–39, 1992.
2. Donald Ballou, Richard Wang, Harold Pazer, and Giri Kumar Tayi. Modeling information manufacturing systems to determine information product quality. *Management Science*, 44(4):462–484, 1998.
3. Arosha K. Bandara, Antonis C. Kakas, Emil C. Lupu, and Alessandra Russo. Using argumentation logic for firewall configuration management. In *Integrated Network Management, IM 2009. 11th IFIP/IEEE International Symposium on Integrated Network Management*, pages 180–187, 2009.
4. Jennifer Bayuk. Data-centric security. *Computer Fraud & Security*, 2009(3):7–11, 2009.
5. Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.*, 93:63–101, 1997.
6. Michael L. Brodie. Data quality in information systems. *Information & Management*, 3(6):245–258, 1980.
7. Hsinchun Chen, Roger HL Chiang, and Veda C. Storey. Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4):1165–1188, 2012.
8. Robert Craven, Jorge Lobo, Jiefei Ma, Alessandra Russo, Emil C. Lupu, and Arosha K. Bandara. Expressive policy analysis with enhanced system dynamicity. In *Proceedings of the 2009 ACM Symposium on Information, Computer and Communications Security, ASIACCS*, pages 239–250, 2009.
9. Thomas H. Davenport and Laurence Prusak. *Information ecology: Mastering the information and knowledge environment*. Oxford University Press on Demand, 1997.
10. Dursun Delen and Haluk Demirkan. Data, information and analytics as services. *Decision Support Systems*, 55(1):359–363, 2013.

11. Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.
12. David F. Ferraiolo and D. Richard Kuhn. Role-based access controls. In *15th National Computer Security Conference*, 1992.
13. Christopher Fox, Anany Levitin, and Thomas Redman. The notion of data and its quality dimensions. *Information processing & management*, 30(1):9–19, 1994.
14. Gerard George, Martine R. Haas, and Alex Pentland. Big data and management. *Academy of Management Journal*, 57(2):321–326, 2014.
15. Michael Gertz and Sushil Jajodia. *Handbook of database security: applications and trends*. Springer Science & Business Media, 2007.
16. Benjamin T. Hazen, Christopher A. Boone, Jeremy D. Ezell, and L. Allison Jones-Farmer. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154:72–80, 2014.
17. Manuel Hilty, Alexander Pretschner, David A. Basin, Christian Schaefer, and Thomas Walter. A policy language for distributed usage control. In *Computer Security - ESORICS*, pages 531–546, 2007.
18. Y. Huh, F. Keller, T. Redman, and A. Watkins. Data quality. *Information and Software Technology*, 32(8):559–565, 1990.
19. L. Jones-Farmer, Jeremy D. Ezell, and Benjamin T. Hazen. Applying control chart methods to enhance data quality. *Technometrics*, 56(1):29–41, 01/02 2014. doi: 10.1080/00401706.2013.804437.
20. Beverly K. Kahn, Diane M. Strong, and Richard Y. Wang. Information quality benchmarks: product and service performance. *Communications of the ACM*, 45(4):184–192, 2002.
21. Antonis Kakas and Pavlos Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '03, pages 883–890, 2003.
22. Antonis C. Kakas, Robert A. Kowalski, and Francesca Toni. Abductive logic programming. *J. Log. Comput.*, 2(6):719–770, 1992.
23. Erisa Karafili, Antonis C. Kakas, Nikolaos I. Spanoudakis, and Emil C. Lupu. Argumentation-based security for social good. *CoRR*, abs/1705.00732, 2017.
24. Erisa Karafili and Emil C. Lupu. Enabling data sharing in contextual environments: Policy representation and analysis. In *Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies, SACMAT 2017, Indianapolis, IN, USA, June 21-23, 2017*, pages 231–238, 2017.
25. Erisa Karafili, Hanne Riis Nielson, and Flemming Nielson. How to trust the re-use of data. In *Security and Trust Management - 11th International Workshop, STM*, pages 72–88. Springer, 2015.
26. Günter Karjoth and Matthias Schunter. A privacy policy model for enterprises. In *Proceedings 15th IEEE Computer Security Foundations Workshop (CSFW-15)*, pages 271–281, 2002.
27. Günter Karjoth, Matthias Schunter, and Michael Waidner. Platform for enterprise privacy practices: Privacy-enabled management of customer data. In Roger Dingledine and Paul Syverson, editors, *Privacy Enhancing Technologies: Second International Workshop, (PET)*, pages 69–84, 2003.
28. Charlie Kaufman, Radia Perlman, and Mike Speciner. *Network Security: Private Communication in a Public World, Second Edition*. Prentice Hall Press, Upper Saddle River, NJ, USA, second edition, 2002.

29. Florian Kelbert and Alexander Pretschner. Data usage control enforcement in distributed systems. In *Third ACM Conference on Data and Application Security and Privacy, CODASPY'13*, pages 71–82, 2013.
30. Florian Kelbert and Alexander Pretschner. A fully decentralized data usage control enforcement infrastructure. In *Applied Cryptography and Network Security - 13th International Conference, ACNS*, pages 409–430, 2015.
31. Young-Jin Kim, Marina Thottan, Vladimir Kolesnikov, and Wonsuck Lee. A secure decentralized data-centric information infrastructure for smart grid. *IEEE Communications Magazine*, 48(11):58–65, 2010.
32. Robert A. Kowalski and Marek J. Sergot. A logic-based calculus of events. *New Generation Comput.*, 4(1):67–95, 1986.
33. Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins, and Nina Kruschwitz. Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 62(4), 2013.
34. Aliaksandr Lazouski, Fabio Martinelli, and Paolo Mori. A prototype for enforcing usage control policies based on XACML. In *Trust, Privacy and Security in Digital Business - 9th International Conference, TrustBus. Proceedings*, pages 79–92, 2012.
35. Yang W. Lee, Diane M. Strong, Beverly K. Kahn, and Richard Y. Wang. Aimq: a methodology for information quality assessment. *Information & Management*, 40(2):133–146, 12 2002.
36. Mark Lycett. "Datafication": Making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4):381–386, 2013.
37. Stuart E. Madnick, Richard Y. Wang, Yang W. Lee, and Hongwei Zhu. Overview and framework for data and information quality research. *Journal of Data and Information Quality (JDIQ)*, 1(1):2, 2009.
38. S. T. March and A. R. Hevner. Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, 43(3):1031–1043, 2007.
39. Ilaria Matteucci, Marinella Petrocchi, and Marco Luca Sbodio. CNL4DSA: a controlled natural language for data sharing agreements. In *Proceedings of ACM Symposium on Applied Computing (SAC)*, pages 616–620, 2010.
40. Andrew McAfee, Erik Brynjolfsson, Thomas H. Davenport, DJ Patil, and Dominic Barton. Big data: The management revolution. *Harvard Business Review*, 90(10):61–67, 2012.
41. Deepa Mishra, Angappa Gunasekaran, Thanos Papadopoulos, and Stephen J. Childe. Big data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research*, pages 1–24, 2016.
42. Marco Casassa Mont and Siani Pearson. Sticky policies: An approach for managing privacy across multiple parties. *Computer*, 44:60–68, 2011.
43. Marco Casassa Mont, Siani Pearson, and Pete Bramhall. Towards accountable management of identity and privacy: sticky policies and enforceable tracing services. In *14th International Workshop on Database and Expert Systems Applications, 2003. Proceedings.*, pages 377–382, 2003.
44. Irene CL Ng, Jason Williams, and Andy Neely. *Outcome-based contracting: changing the boundaries of B2B customer relationships: an executive briefing.* Advanced Institute of Management Research, 2009.
45. Bert Van Nuffelen and Antonis C. Kakas. A-system: Declarative programming with abduction. In *Logic Programming and Nonmonotonic Reasoning, 6th International Conference, LPNMR, Proceedings*, pages 393–396, 2001.
46. David Opresnik and Marco Taisch. The value of big data in servitization. *International Journal of Production Economics*, 165(C):174–184, 2015.

47. Jaehong Park and Ravi S. Sandhu. Towards usage control models: beyond traditional access control. In *SACMAT*, pages 57–64, 2002.
48. Jaehong Park and Ravi S. Sandhu. The ucon_{abc} usage control model. *ACM Trans. Inf. Syst. Secur.*, 7(1):128–174, 2004.
49. Siani Pearson, Marco Casassa Mont, Liqun Chen, and Archie Reed. End-to-end policy-based encryption and management of data in the cloud. In *2011 IEEE Third International Conference on Cloud Computing Technology and Science*, pages 764–771, 2011.
50. Milan Petkovic, Davide Prandi, and Nicola Zannone. Purpose control: Did you process the data for the intended purpose? In *Proceedings of Secure Data Management - 8th VLDB Workshop, SDM*, pages 145–168, 2011.
51. Leo L. Pipino, Yang W. Lee, and Richard Y. Wang. Data quality assessment. *Communications of the ACM*, 45(4):211–218, 2002.
52. Alexander Pretschner, Manuel Hilty, and David A. Basin. Distributed usage control. *Communications of the ACM*, 49(9):39–44, 2006.
53. Alexander Pretschner, Manuel Hilty, David A. Basin, Christian Schaefer, and Thomas Walter. Mechanisms for usage control. In *Proceedings of the 2008 ACM Symposium on Information, Computer and Communications Security, ASIACCS*, pages 240–244, 2008.
54. Thomas C. Redman. The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2):79–82, 1998.
55. Thomas C. Redman and A. Blanton. *Data quality for the information age*. Artech House, Inc., 1997.
56. Boaz Ronen and Israel Spiegler. Information as inventory: a new conceptual view. *Information & management*, 21(4):239–247, 1991.
57. Nikolaos I. Spanoudakis, Antonis C. Kakas, and Pavlos Moraitis. Gorgias-b: Argumentation in practice. In *Computational Models of Argument - Proceedings of COMMA*, pages 477–478, 2016.
58. Diane M. Strong, Yang W. Lee, and Richard Y. Wang. Data quality in context. *Communications of the ACM*, 40(5):103–110, 1997.
59. Vipin Swarup, Len Seligman, and Arnon Rosenthal. Specifying data sharing agreements. In *7th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY)*, pages 157–162, 2006.
60. Slim Trabelsi and Jakub Sendor. Sticky policies for data control in the cloud. In *Proceedings of the 2012 Tenth Annual International Conference on Privacy, Security and Trust*, PST '12, pages 75–80. IEEE Computer Society, 2012.
61. Matthew A. Waller and Stanley E. Fawcett. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2):77–84, 2013.
62. Samuel Fosso Wamba, Shahriar Akter, Andrew Edwards, Geoffrey Chopin, and Denis Gnanzou. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165:234–246, 2015.
63. Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou. Privacy-preserving public auditing for data storage security in cloud computing. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, 2010.
64. Gang Wang, Angappa Gunasekaran, Eric WT Ngai, and Thanos Papadopoulos. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176:98–110, 2016.

65. Richard Y. Wang. A product perspective on total data quality management. *Communications of the ACM*, 41(2):58–65, 1998.
66. Richard Y. Wang, Martin P. Reddy, and Henry B. Kon. Toward quality data: An attribute-based approach. *Decision Support Systems*, 13(3):349–372, 1995.
67. Richard Y. Wang and Diane M. Strong. Beyond accuracy: What data quality means to data consumers. *J. Manage. Inf. Syst.*, 12(4):5–33, 1996.
68. Richard Y. Wang and Diane M. Strong. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4):5–33, 1996.
69. Karl E. Weick. *Sensemaking in organizations*, volume 3. Sage, 1995.
70. Xinwen Zhang, Francesco Parisi-Presicce, Ravi Sandhu, and Jaehong Park. Formal model and policy specification of usage control. *ACM Trans. Inf. Syst. Secur.*, 8:351–387, 2005.
71. Wenchao Zhou, Micah Sherr, William R. Marczak, Zhuoyao Zhang, Tao Tao, Boon Thau Loo, and Insup Lee. Towards a data-centric view of cloud security. In *Proceedings of the Second International Workshop on Cloud Data Management*, CloudDB '10, pages 25–32, 2010.