Article

Multilevel Privacy Assurance Evaluation of Healthcare Metadata

Syeda Amna Sohail * , Faiza Allah Bukhsh and Maurice van Keulen

Data Management and Biometrics (DMB) Department, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Zilverling, P.O. Box 217, AE 7500 Enschede, The Netherlands; f.a.bukhsh@utwente.nl (F.A.B.); m.vankeulen@utwente.nl (M.v.K.)

* Correspondence: s.a.sohail@utwente.nl; Tel.: +31-53-489-9788

Abstract: Healthcare providers are legally bound to ensure the privacy preservation of healthcare metadata. Usually, privacy concerning research focuses on providing technical and inter-/intra-organizational solutions in a fragmented manner. In this wake, an overarching evaluation of the fundamental (technical, organizational, and third-party) privacy-preserving measures in healthcare metadata handling is missing. Thus, this research work provides a multilevel privacy assurance evaluation of privacy-preserving measures of the Dutch healthcare metadata landscape. The normative and empirical evaluation comprises the content analysis and process mining discovery and conformance checking techniques using real-world healthcare datasets. For clarity, we illustrate our evaluation findings using conceptual modeling frameworks, namely e^3 -value modeling and REA ontology. The conceptual modeling frameworks highlight the financial aspect of metadata share with a clear description of vital stakeholders, their mutual interactions, and respective exchange of information resources. The frameworks are further verified using experts' opinions. Based on our empirical and normative evaluations, we provide the multilevel privacy assurance evaluation with a level of privacy increase and decrease. Furthermore, we verify that the privacy utility trade-off is crucial in shaping privacy increase/decrease because data utility in healthcare is vital for efficient, effective healthcare services and the financial facilitation of healthcare enterprises.

Keywords: healthcare; privacy; privacy–utility tradeoff; process mining; REA ontology; e^3 -value modeling

Citation: Sohail, S.A.; Bukhsh, F.A.; van Keulen, M. Multilevel Privacy Assurance Evaluation of Healthcare Metadata. *Appl. Sci.* **2021**, *11*, 10686. <https://doi.org/10.3390/app112210686>

Academic Editor: Gianluca Lax

Received: 1 October 2021

Accepted: 6 November 2021

Published: 12 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Big Data Analytics (BDA) of Electronic Health Record (EHR) is a prerequisite in advancing efficient and effective healthcare and clinical research [1]. BDA is the application of advanced analytical methods for big data streams to extract useful value [2]. For example, Big Data Analytics (BDA) is significant in understanding, rationalizing, and utilizing healthcare data for multiple purposes [3]. However, while performing BDA, healthcare providers must avert from data-related ethical issues such as biased decision making, unwanted disclosures, and inaccuracies. In this wake, the healthcare providers ascribe their healthcare data analytics practices to responsible data science [4]. Amongst others, privacy preservation of the sensitive healthcare metadata is one such practice. Here, metadata is data about data or structured data, such as structured patients' data accompanying lab measurements or images.

Privacy, in a nutshell, comprises an individual's autonomous decision making and direct or indirect control over his/her personal information [5]. Across nation-states, the privacy preservation of patients' Personally Identifiable Information (PII) is legally ensured using some fundamental multilevel privacy-preserving measures [6–9]. The fundamental multilevel privacy-preserving measures include the following:

- The process of privacy-preserving measures initiates with the intake of patients' Informed Consent (IC), concerning patients' information-share, at the onset of their data collection [10–13].;
- Privacy by Design (PbD) measures involve the privacy incorporation into the design specifications of technologies [14,15]. PbD begins with embedding prime privacy-preserving principles into designing, operating, and managing information processing technologies and systems. PbD then allows the elaboration of those principles as default standards;
- Privacy by policy includes measures where healthcare providers are accountable to and seek accreditation from global/pan-European supra-organizations and local regulators [7–9,16]. Information Security Management Systems are evaluated for privacy preservation using policy measures [17];
- The privacy-preserving measures for third-party data processing include the tasks for data controller and processor as per GDPR (General Data Protection Regulation) [18].

Multilevel privacy assurance evaluation is a term coined to attribute the normative and empirical evaluation of (aforementioned) multilevel privacy measures. The focus is on Dutch healthcare metadata handling. Metadata handling includes data collection, storage, processing, and recovery/retention [19]. Here, normative means “should be”, and empirical means “as is” state of affairs [20]. The multilevel privacy assurance evaluation of privacy-preserving measures is divisible into three levels of Dutch healthcare metadata handling that are enumerated below:

1. Individual level is from the perspective of the individual patient, where the patient's PII is collected, after his or her Informed Consent (IC), at the onset of data collection;
2. Inter/intra-organizational level includes the healthcare information-share amongst horizontally and vertically distributed organizations. The horizontally distributed organizations include the locally located caregivers within a healthcare organization (such as surgeons or radiologists within a hospital or lab professionals within a diagnostic lab). Whereas the vertically distributed organizations include the remotely located caregivers (such as General Practitioners (GPs), hospital, and diagnostic labs) who remotely perform patients' information-share;
3. Third-party metadata processing level includes the patients' information-share with a Trusted Third Party (TTP), who performs data processing for further value extraction of healthcare metadata. TTP usually remotely conducts healthcare research.

Figure 1 shows that patients' Electronic Health Record (EHR) is crucial for healthcare providers for quality BDA [1]. The quality BDA directly depends upon the provenance records, i.e., the genuine documentation of the metadata pipeline. The metadata as a final output are then stored in data warehouses in either remote or onsite servers for metadata retrieval and interoperability. The data analysts encounter ethical issues at each phase of the metadata pipeline: from data collection to metadata formation, storage, and interoperability [2]. To counter ethical issues, data analysts abide by principles such as FACT (i.e., Fair, Accessible, Confidential, and Transparent) [4,21]. Conversely, to broaden the scope of healthcare metadata utility, data analysts have to adapt to FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Thus, the ultimate goal determines the following path while handling metadata.

The privacy preservation of sensitive healthcare metadata is legally desired at each phase of the metadata pipeline. In this wake, Privacy by Design, privacy by policy (via ISMS), and informed consent are fundamental privacy-preserving measures to ensure the ethical usage of healthcare metadata. However, the concerned authorities encounter the Privacy Utility Tradeoff (PUT) while performing fundamental privacy-preserving measures. PUT is the performance impairment of data utility in ensuring data privacy [20]. This research work focuses on exploring this problem area. Figure 1 shows the significance of responsible data science and its direct influence upon EHR and BDA. The figure portrays how privacy is an essential measure of responsible data science for the stakeholders and

affectees alike. The research aims to highlight PUT's influence across multilevel privacy assurance measures.

Figure 1. Description of the Problem Area.

1.1. State of the Art

Privacy as a fundamental human right: privacy is a fundamental human right and is legally protected at global [22], European [23], and national levels [7]. However, the idea of privacy is sometimes elusive and far reaching. From the individual's rights to be left alone and to control personal information share, to organizations' responsibilities concerning the collection, execution, disclosure, and retention of individuals' PII, all come under the umbrella term of 'privacy' [24]. In a nutshell, privacy is an individual's autonomous decision making and direct/indirect control over personal information share [5].

BDA for Better Data Utility and Ethical Concerns: With the growing dependency upon BDA, the individual's autonomy and decision making is now shifting from individuals to technology for better data utility [19]. Data utility directly depends upon the correct use of big data by implying apt data techniques to deliver efficient/effective services/goods [3]. Better data utility is also significant for financial stability as a majority of the western European and Scandinavian countries allocate a substantial share of approximately 10% of their GDP to healthcare [25]. However, there are also concerns regarding the fair use of data [2,19]. For BDA, healthcare providers enumerated some generally prioritized values while dealing with healthcare data at the organizational level [26]. Among those ten enumerated values, four are directly related to patients' information security. In comparison, the rest of the six values regard the uninterrupted efficient healthcare data sharing for better data utility. Patients' information security values include personalized healthcare, patients' value care risks, privacy protection, and transparency, where privacy preservation is the least-valued organizational value for BDA.

Patient-Centric, Value-Based Healthcare Model: Healthcare providers moved from a former, volume-based fee-reimbursement healthcare model to a value-based, patient-centric healthcare model [27]. In a value-based healthcare model, physical first-order value transactions between the value actors are the same as volume-based healthcare. However, the difference lies in qualitative intrinsic values, which prioritize patients' satisfaction along with patients' health improvement [17]. The shift marks the transformation from an exclusive monetary gain-oriented business model to a more intrinsic value-based (i.e., privacy and data security) value-modeling for consistent, trust-based future interactions [28]. Trust is an integral aspect of a value model, and it has to persist between the value actors throughout the healthcare service for healthy outcomes. The patient-centric value model also seeks shared decision making from patients' first meeting with the GP to the diagnosis, treatments, and better healthcare outcomes. In this wake, prioritized healthcare is mutually decided between the caregiver and the patient. Shared decision making is significant for patients' engagement, adherence, cost-effective decision making, and over-

all satisfaction [29]. Value-based healthcare is practiced mainly for more cost-effective, quality-assuring, and more patient-engaging health outcomes for the benefit of related stakeholders ranging from healthcare providers to the patient himself [30].

BDA and Informed Consent at the Expense of Privacy: Consumerism is defined as the “advocacy of the rights and interests of the consumers” [31]. Consumerism in healthcare advocates the patients’ empowerment and their participation in decision-making regarding their healthcare. From a ‘consumerism’ perspective, privacy is protected personal information-use as per data subjects’ expectations. This involves the idea of Informed Consent (IC) of the data subjects and the idea of patients’ respective notice and choice according to the privacy by policy indicators in the healthcare domain [16]. In this wake, bigger data and better algorithms have gradually mitigated the significance of privacy breaches of individuals in favor of their informed consent [2]. IC intake takes place at the onset of patients’ PII collection. However, not only at the onset of data collection, rather each step in the data pipeline causes ethical issues, i.e., inaccuracy, inequality, non-transparency, and intrusion [21]. If emerging technologies are to be classified concerning their ethical issues, each one would have the prime issue of privacy protection [32].

Privacy-by-design parameters aim to technically integrate the privacy indicators into the organization’s Information Systems [14]. Privacy by design is system based and incorporates ISMS standard parameters in alignment with the organization’s objectives and goals. Usually, PbD takes place in three steps: namely, minimization, enforcement, and transparency [15]. On the other hand, privacy by policy for organizations and individuals is the application of regulations, laws, policies, and processes by which personal information is managed [16]. Privacy by policy is mainly related to privacy-securing data processing of data subjects. Its byproduct is the Information Security Management System (ISMS), with availability, integrity, confidentiality, and accountability as its standard indicators. The first two indicators encompass “patients security”, and the remaining two involve “patients privacy”, which collectively come under the tag of Information Security [17].

Privacy Utility Tradeoff, PUT: To ensure the responsible healthcare metadata handling, the main concern, amongst others, is the Privacy Utility Tradeoff (PUT) [3,33]. PUT appears when the privacy-preserving measures (tools and techniques) start hampering the data utility after a point. Data utility is the precision of the analytical data results [34]. In healthcare, the PUT raises graver repercussions because of the involvement of highly sensitive Personally Identifiable Information (PII) on the one hand and the efficiency and effectiveness of healthcare services on the other [35]. Therefore, research studies focus on multiple domains separately. For example, privacy and IoT [36–38], privacy in AI, and machine learning [39,40] privacy in Process Mining [41–44]. This targeted attention lagged in paying concerted attention to the overall context of the issue.

Privacy and Third-Party Data Processing using Process Mining: The privacy concerns ameliorate when healthcare metadata are shared with a third party (data processor) for further data evaluations using Process Mining (PM) tools and techniques. Process Mining (PM) is an emerging discipline that bridges the gap between data mining and business process management. PM involves process optimization and includes business deviations and bottlenecks finding by utilizing its discovery, enhancement, and conformance checking techniques [45–48]. In PM, the event logs are extracted from an organization’s Information System (IS) for an insightful evaluation of actual business proceedings [41,44]. PM is not inherently privacy preserving and seeks external privacy preserving measures.

Usefulness of Process Mining during the COVID-19 pandemic: PM as an online third-party metadata processing (collection of) technique(s) proved to be highly useful during the COVID-19 pandemic across domains. For example, by evaluating the metadata of Covid 19 patients in Uniklinik Aachen hospital, PM could facilitate the hospital in reconstructing their ICU model for cost and time-efficient treatments [49]. In the logistics domain, PM, using virtual desktop infrastructure, ensured productive remote controller operations while keeping intact the integrity and confidentiality of relevant business information [50]. PM is

also highly applicable in automating the business processes in the post-COVID economy with the help of artificial intelligence and robotic process automation [51].

PM and privacy gaps: In PM, to maintain the performance of data analytics, the event logs are kept un-anonymized. This privacy lapse allows the adversary to take advantage of the readily available PII of the data subjects. For this reason, many healthcare providers are reluctant to share sensitive event logs with third-party data processors outside the premises of their organization. In this regard, the *Privacy-Preserving Process Mining (PPPM)*: The proposed Privacy-Preserving Process Mining (PPPM) framework seeks further privacy-preservation by adding noise to pseudonymized healthcare dataset [43]. For PPPM, the metadata with a detailed description of added noise type, quantity, and purpose must be shared with a Trusted Third Party (TTP) [41,43,52–54].

Differential Privacy (DP): New techniques and tools are under consideration to maximize data utility while intrinsically integrating privacy in healthcare analytics. DP is well acclaimed for striking a fair balance between privacy preservation and data utility across multiple domains [55–61]. DP is achieved by proportionately adding noise to the event log. The noise is added to obscure the reidentification of data subjects while maintaining the accuracy/correctness of data analytics [55,56]. In PPPM, the protection models based on the Differential Privacy notion provide means to thwart potential privacy leakages in event logs [41].

Noise Addition and PUT in Healthcare Metadata and Process Mining: Concerning PUT, the healthcare providers avoid noise-addition and other such data-alteration methods to ascertain overall healthcare performance [62]. For example, noise addition (for privacy preservation of the datasets) is avoided in the clinical practice-trials database to maintain data utility). Similarly, in Process Mining (PM), disproportionately added noise to datasets can hamper the usefulness of the original datasets (i.e., data utility) [43] causing PUT. The PUT dilemma motivated us to use real-world (healthcare) datasets for the evaluation of the (publicly available) ProM tool's [63] noise-adding plugins. We evaluate the plugins to precisely judge them for their data utility using PM conformance checking indicators. The conformance checking indicators include fitness, precision, and complexity. The privacy-preserving third-party data processing evaluation allowed finding the influence of PUT from the perspective of promtool end-user using publicly available, real-world healthcare event logs.

Considering the importance of privacy assurance and the fears it entails, it is vital to evaluate the current situation as a priority. Only then can improved set standards guarantee privacy protection for all stakeholders with standardized indicators and their respective metrics. Furthermore, value modeling bridges the gap between IT and organizational undertakings. The conceptual modeling value frameworks also highlight the financial aspect of mutual undertakings in the Dutch healthcare landscape.

1.2. Research Aim, Purpose, and Main Conclusions

The aim of this research work is to provide a multilevel privacy assurance evaluation of fundamental privacy-preserving measures for Dutch healthcare metadata and finding privacy increase/decrease at each level. The multilevel privacy assurance evaluation comprises normative and empirical evaluation. The normative evaluation is performed using content analysis. The empirical evaluation is conducted using Process Mining on real-world healthcare event logs. The evaluations' findings are illustrated using, 'Padlock Chain Model' of e^3 -value modeling and the 'Insurance Model' of Resource Event Agent (REA) ontology. The resulting conceptual modeling frameworks highlight the financial aspect with vital stakeholders, their mutual value transactions of data-oriented resources in a confined time. The purpose of this research work is three-fold: firstly, the exploratory and data-analytical research work of multilevel privacy assurance; secondly, the evaluation findings' illustration using e^3 -value models and REA ontology conceptual modeling frameworks and their verification using expert opinion; and thirdly, to find multilevel increase/decrease in privacy for Dutch healthcare metadata and its direct dependency upon Privacy Utility

Trade-off (PUT). The research work is helpful for (academia and industry alike) to collectively rectify the verified privacy gaps that deserve attention and immediate fixation before reaching out for extensive technical and organizational privacy-specific intricacies.

The main conclusions of this research work provide the multilevel privacy assurance evaluation of fundamental privacy-preserving measures of healthcare metadata in the Dutch healthcare landscape. We performed the empirical and normative evaluation and illustrated the findings using e^3 -value modeling and REA ontology. We explored the increase/decrease in privacy at each privacy assurance level. We verify that Privacy Utility Tradeoff plays a vital role in shaping the increase/decrease in privacy for the sake of better data utility. This utility-prone metadata handling primarily ensures uninterrupted healthcare metadata share for patients' efficient/effective healthcare, which ascertains the service-dominant logic of patients' value care. Moreover, this uninterrupted healthcare metadata share predominantly financially facilitates the healthcare enterprises (a fundamental objective of GDPR, 2018). The latter is an essential contribution to this research work.

In this paper, Section 2, namely Materials and Methods, provides an overarching methodology with a data analytical approach. It also incorporates a separate description of the followed methodology for each level of privacy assurance evaluation. Section 3 constitutes the results and findings based on normative and empirical evaluations. Section 4 discusses and concludes the multilevel privacy assurance evaluation. Section 4 also proposes a multilevel privacy assurance curve with direct implications from the PUT in handling sensitive healthcare metadata in Dutch healthcare. The contribution of the research work comes at the end of Section 4.

2. Materials and Methods

The 'refined Process Mining (PM) framework' [42] laid out the Process Mining spectrum and inspired us to follow similar steps in executing this research work. The framework inspired our research approach because the refined PM framework incorporates the controlflow aspect of "How?", the organizational aspect of "Who?", and the metadata aspect of "What?" of the business processes. The framework also highlights the usages and contributions for analyzing the premortem (currently opened cases) and post mortem (already completed historic cases) in event logs. Event logs are the datasets that are extracted directly from an organization's IS for data-driven evaluations. Furthermore, the idea of normative (i.e., de jure) and empirical (i.e., de facto) business models is also described in this framework, where goals are cartography, auditing, and navigation. Cartography is the extraction and improvement of de facto empirical process models deduced from event logs. Auditing considers the de jure (should-be state that is assumed to be correct) business model and compares it against the de facto business process model for improvements' sake. This sometimes improves the de jure model if the empirical model is performing better than expectations. Navigation improves the future based on normative and de facto evaluations with future recommendations.

Process Mining spectrum covers the who, how, and why of a case study (in this case, the healthcare metadata share in the dutch healthcare landscape). However, the normative evaluation of the documents requires text mining of official documents, which are usually not publicly available. For this reason, the normative evaluation in this research work is based on content analysis using study of the literature, official websites, and publicly available official documents. The used methodology with data analytical approach for this research work is shown in Figure 2. For PM-oriented de jure analysis, the PM-based text mining of one of the Dutch diagnostic labs is underway for future navigation of a local diagnostic lab in the Netherlands.

For empirical evaluation, Process Mining (PM) discovery and conformance checking techniques are used. PM is performed using publicly available, i.e., ProM Lite 1.2, ProM 6.9–6.10 [63] tools, and commercially available Disco [64] tool. PM is performed on publicly available, real-world healthcare event logs [65]. Based on the evaluations, the conceptual

modeling frameworks are drawn using Resource Event Agent (REA) ontology [66] and e^3 -value modeling [67]. Furthermore, the expert's opinion further verified the evaluations and conceptual modeling frameworks. The underlying reasons behind choosing the aforementioned conceptual modeling frameworks were:

- To highlight the significant stakeholders, their vital interactions with their mutual exchange of information resources while handling healthcare metadata;
- Moreover, the ontologies and the privacy-aware mutual interactions also accentuate the monetary aspect of the transactions. The latter is an essential contribution to this research work.

The conceptual modeling frameworks are verified with the expert(s) opinion. Finally, an increase/decrease in privacy is demonstrated at each level primarily influenced by the PUT of Dutch healthcare metadata.

As mentioned earlier, multilevel privacy assurance evaluation is divided into three levels. For each level, the overarching methodology is shown in Figure 2. Moreover, the elaborated methodology for each subsequent level is mentioned under each level, namely individual level, inter- and intra-organizational level, and third party data processing level.

Figure 2. Research methodology with data analytical approach.

2.1. Privacy Assurance Evaluation at Individual Level

The first model represents the current state of privacy-protecting metadata sharing between patients and healthcare providers (from the patients' perspective). Based on content analysis, the findings are illustrated using e^3 -value kit on the guidelines of Padlock Chain Model [68] in a Dutch healthcare setting.

Why e^3 -value kit?: e^3 -value model provides with the web of enterprises where value actors are key stakeholders from the "creation, execution, and consumption" of the value objects. Value objects are goods and services that are mutually exchanged between the value actors. Linear relationships emphasize that the value actors undergo value transfers of value objects for a limited time as part of an ideal situation where they trust one another for the timely mutual satisfaction of their individual needs. Thus, trust is a constant in value modeling. Presumably, with the e^3 -value model, it is easier to unravel the optimal value care, which assures patients privacy with uninterrupted pooling of medical data. Moreover, it becomes easier to utilize the e^3 -value model for an otherwise complex healthcare ecosystem. This is because it focuses on linear inflow and outflow of the value transfers within a time frame between key-value actors. The linear temporal and spatial transactions facilitate diagnosing the bottlenecks (if any) in performing effectively and efficiently in a value transaction. Value actors are divisible into four main headings that are following:

1. Principals: value actors who wish to satisfy their prime need;
2. Agent: value actors whom the principal chooses to satisfy his or her prime need;

3. Regulator: value actors who regulate and give accreditations to the agents provided that the latter satisfies the prerequisites of being privacy-preserving entities;
4. Third Party: value actors who assist and facilitate the agents but are not directly contacting the principal.

Similarly, these four main actors perform the following four main (value) activities:

1. Front-end activity: principal undergoes this activity;
2. Counter response: agent responds with the counter response against the front-end activity;
3. Regulation and giving accreditations: regulators and international and national standard organizations regulate and give accreditations;
4. Back-end activity: third party performs back-end activity.

The aim is to analyze privacy-preserving measures with simultaneous, uninterrupted, pooling of healthcare data across different domains of healthcare providers.

At the individual level, Informed Consent (IC) entails patients' explicit, well-informed, specific, and autonomous consent for his or her healthcare metadata handling. The metadata handling includes data collection, metadata formulation, storage, and processing. The IC must be demonstrated in an easily accessible and understandable manner, with patients' having the right of erasure at any point (GDPR, article 7). For the individual level and specifically for the IC part, no empirical evaluations are conducted because the privacy limitations do not rely upon the procedure of taking informed consent, rather on the follow-up procedures of the metadata pipeline. In addition, IC poses limitations, specifically against clinical research and third-party data processing. Limitations range from data collection to data anonymization/pseudonymization to metadata storage, sharing, and retrieval.

2.2. Privacy Assurance Evaluation at Inter- and Intra-Organizational Level

e³-value-modeling: In this model, the focus is drawn between the privacy assuring bilateral relationship between a diagnostic lab (biobank and biodepository) as a healthcare provider with other key stakeholders in the Dutch healthcare backdrop.

Laboratories (biobanks and biorepositories) play a pivotal role in facilitating healthcare data sharing. Samples and data collected from them pave the way for valuable clinical research. Efficient sharing and pooling of this data is an essential prerequisite for advancements in biomedical science [69]. In this respect, apprehension resides regarding "where data goes", "by whom it is used", and "for what purpose", which, so far, had not been dealt with aptly [70]. To simplify the understanding of this rather complex value network, the Padlock Chain Model provides a linear and easily comprehensible model to start with. This model represents the lab's bilateral relationship with other key-value actors in privacy-preserving healthcare data sharing.

Process Mining and Resource Event Agent (REA) ontology: We discovered patients' un-anonymized PII share using Process Mining on Dutch healthcare event logs. Patients' un-anonymized information share occurs between the horizontally distributed caregivers (who are located within a hospital but belong to different suborganizations within the hospital). We located patients' sensitive information-share with their initiating and ending timestamps for each event. For example, one event belongs to a case (i.e., individual), an activity, and a timestamp. After the evaluations, we further verified our results with an IT expert affiliated with the Dutch hospital and diagnostic labs in Overijssel province (Netherlands). The expert substantiated that our results show the actual data share status quo in the Dutch healthcare landscape. The results also apply to the metadata shared across vertically distributed caregivers (hospitals, labs, pharmacies, and GPs). Furthermore, we extended the Insurance Model (IM) of REA ontology [66] highlighting the key-value agents, their prime resource interactions, and mutual value gain/loss. Additionally, we explicated privacy as a Materialized Claim in the Dutch healthcare landscape using REA.

2.3. Privacy Assurance Evaluation at Third Party Data Processing Level

For third-party data processing, the focus is only kept within the PM domain. Primarily because the privacy concerns relating to the Internet of Things (IoT) with recent healthcare wearables, Artificial Intelligence (AI), and clinical research are already well renowned [71,72]. Moreover, the concerns increase with the availability of disease-specific online EHR with sensitive information of patients, nurses, and specialists to reduce mobility for both patients and home care providers [73]. This research focuses on PM data analytics by third-party data processors who work remotely instead of working on the field within a healthcare organization's premises. Unfortunately, the former experiences a trust deficit, whereas the latter is often easily trusted for the benefit of both. For example, MRON, Radiology East Netherlands [74] employees are located (along with their radiology machines) and are working within the premises of Medisch Spectrum Twente (Enschede and Oldenzaal) and Ziekenhuisgroep Twente (Almelo and Hengelo) hospitals in Overijssel province of Netherlands, where big data analytics is an integral part of their routine proceedings for efficient/effective healthcare services.

Similarly, in Overijssel, private diagnostic labs (biobanks) are located within the premises of local hospitals [75]. Medlon is one such example that is a member of Unilab (a prominent European tycoon in lab, imaging, and pathology diagnostics with labs majorly across Europe and in a few parts globally). Medlon is working with its sister firms in the pharmaceutical manufacturing industry as well. Biobanks deal with DSE (Data and Sample exchange) within and across healthcare providers and play a vital role in improving healthcare services and keeping track of treatments and respective patient responses. Biosamples and biodata are then kept safe for further research [13]. The Process Mining real-world data-oriented process analytics are used to explore the privacy assurance evaluation level.

Datasets and Datasource: The Sepsis event-log is a "Sepsis" patients', publicly available healthcare event log, from an Electronic Health Record (EHR) of a local Dutch hospital [65]. The event log has up to 16 activities for 1050 cases and 15,214 events. 'Hospital billing' is another publicly available event log extracted from a regional hospital IS with the billing of healthcare packages of the patients with 18 activities, 100,000 cases, and 451,359 events [65]. Another publicly available (healthcare) event log (i.e., Urineweginfectie (UWIncasus) logboek) [65] was used. However, it gave the lowest utility because of its sensitive nature (presumably because of a biosample that requires exclusive privacy preservation). It, therefore, was not added below in the results of the experiments.

3. Multilevel Privacy Assurance Evaluation: Results and Evaluation

This section provides a concise and precise description of the experimental results, their interpretation, and the experimental conclusions drawn with the discussion.

3.1. At Individual Level: Results and Evaluation

The e³-value modeling represents a value network to match information and organizational requirements while sharing sensitive healthcare data. It is based on Padlock Chain Model, focusing on privacy preservation as per respective subdomains privacy-preserving means.

3.1.1. Results

As an element of consumerism and service-dominant logic [76], the value chain starts with the healthcare consumers, i.e., the patients. Thus, the healthcare consumers, as the market segment of patients, are the principal. The principal utilizes the postulates of privacy by policy, privacy by design/architecture, and the patient's informed consent; see Figure 3. Principal at the onset of ailment delegates the duty to the agents (i.e., healthcare provider). In return, the principal provides his or her PII to agents while interacting with them. In an ideal patient value-care scenario, if a principal offers his or her sensitive personal info to other value actors, the privacy is a common (intrinsic) factor that he or she expects in return alongside other tangible healthcare outcomes.

Agent, on the other hand, performs the assigned duty in a confined period called contract [77]. In this wake, the outcome-oriented contracts are most efficient as they co-align the agents' priorities with that of the principal regarding risk-sharing. This co-alignment originates rewards for the agents by providing value care to principal [77]. For example, the Statement of Applicability (ISO 27001 Clause 6.1.3 d) links the risk assessment and treatment and the implementation of information security. Similarly, a transparent information system also serves the principal as it reveals the agents' undertakings from the beginning of the contract [77].

According to the international (ISO27001) and the Dutch (NEN7510) standards for Information Security Management Systems (ISMS), the Netherlands were certified for Transparency in Healthcare (TiH) On 7 September 2015. Transparency in Healthcare (TiH) meets the highest standards for data security and regulatory compliance [78]. However, some Dutch healthcare providers are in the lead concerning TiH than the rest. For example, hospitals usually make a transparency window with the patients for patients' satisfaction, but the rest of the caregivers, including GPs, lag far behind in this.

Figure 3. Privacy assurance evaluation at individual-level.

Regulators are divisible into two types. One type of regulator gives accreditations based on global and national standards (i.e., ISO and NEN) to the healthcare providers based on their compliant behaviors. The other type comprises independent agency and regulatory authority like Inspectie Gezondheidszorg jeugd (IGJ) and Nederlandse ZorgAutoriteit (NZA), respectively. IGJ performs the oversight tasks, whereas NZA remotely regulates and facilitates competition in the highly disintegrated Dutch healthcare market [9]. This type of regulator has powers to ask for compensations based on noncompliant behavior [12].

The third party is the value actor who is not directly involved with the principal but facilitates the agent in efficiently executing value transactions with the principal, for example, the pharmaceutical manufacturers, system-based machine vendors to healthcare providers. The value transactions take place as core and counter-value activities. Core activities are the initiating activities of value actors. By contrast, counter activities are

response-based activities. The value activities with the regulators are proof and counter activities. The proof includes the proof of performance to the regulator by healthcare providers concerning privacy preservation. The counter activity includes the confidence in healthcare providers' information security from the regulator. During value transfers, for patients, there is an outflow of data-carrying personal sensitive info of the patient to the domain-specific healthcare providers such as GPs, pharmacy, hospital+EHBO, diagnostic lab (biobank and biorepository), and health insurance company. There is an inflow of healthcare services to the patients. Additionally, there is an expected inflow of privacy as a (subjective) value object to healthcare consumers.

Patients expect privacy preservation of their PII from each healthcare provider that is directly/indirectly involved from data collection to metadata sharing and storage. It is to keep in mind that when data securing sensitive personal information is collected, stored, or accessed, different data protection requirements are met in different phases of data handling. Similarly, different privacy issues arise at different stages of data handling and are handled differently as per technical/organizational objectives and goals. As e³-value modeling focuses more on the stakeholders and their respective priorities, similarly value-based healthcare models prioritize patients for the benefit of all. Thus, it exhibits that "Who" is offering "what", of value to whom and expects "what" in return.

3.1.2. Evaluation

Privacy as (Subjective) Value Object: Generally, it is a well-ingrained notion that e³-value modeling better demonstrates the physical/tangible value transactions. The idea is that the e³-value modeling better wholly demonstrates tangible/physical value objects instead of qualitative value transactions such as privacy, security, transparency, autonomy, and confidentiality. As the latter are all respective terms and depend mainly upon the purposes and requirements of the stakeholders. The e³-value model represents privacy from a subjective value to a (subjective) value object as it satisfies all four attributes [68] to become a value object. The model explores privacy preservation as per respective subdomains privacy-preserving means. Finally, privacy transforms from an intrinsic value to a (subjective) value object that requires assurance for patients' value care.

In this wake, the privacy by policy and PbD serve as privacy-preserving measures when a patient comes in contact with a healthcare provider. The metadata privacy preservation initiates with the patients' PII collection after the intake of patients' Informed Consent (IC). Here, using the padlock chain model of e³-value modeling, patients' privacy is received as a (subjective) value object along with other tangible value activities such as quality healthcare. Quality healthcare is time- and cost-effective treatment of patients after the intake of their IC, while privacy is verified as a subjective transaction that varies according to each value actor's ontological requirements/priorities. Other tangible transactions (services) are physical. By contrast, privacy is perceivable as a (subjective) value object as it satisfies four attributes of being a value object in the padlock chain model. Below are the attributes with respective logic that proves privacy to be a (subjective) value object:

1. Business goal: Service-Dominant Logic preaches that patients' autonomous decision making and their right over the control of their PII sharing must be prioritized;
2. Proof of Performance (POP): healthcare providers give their proof of performance to local oversight and regulatory authorities, such as IGJ and NZA, for patients' privacy preservation;
3. Accreditation: The healthcare providers are accredited by standard authorities such as ISO and NEN;
4. Exchanged as a resource: privacy is exchanged as a resource from healthcare providers to patients and concerned authorities alike.

Currently, privacy satisfies the above attributes to be reckoned as a value object but only subjectively—subjective, because the privacy assurance evaluation by accrediting and regulatory bodies is still nominal that relies on named indicators. Nominal evaluation allows indicator-based qualitative assessment and lags in ordinal, interval, and ratio-

based measurements. Ordinal, interval, and ratio-based measurements include evaluation measures that characterize evaluation measures that assert the set order of priority, proportion, and quantitative measurements of named indicators. For quantitative evaluation, the healthcare providers' implementation, enforcement, and transparency (privacy by design) should be evaluated at the backdrop of the ISMS indicators, i.e., availability, integrity, confidentiality, and accountability.

Informed Consent (IC): To transform the former, subjective healthcare metadata handling by healthcare providers into an objective pan-European one, the EU introduced General Data Protection Regulation, GDPR, article 7. In this wake, a transition of IC is perceivable that ranges from pre-GDPR time to post-GDPR time to GDPR-future expectations-oriented time.

In the pre-GDPR era, accuracy and context preservation of the healthcare metadata was of prime significance for better data evaluations. Information security, the openness of healthcare metadata, and ethical issues relating to it were already realized across academia and industry alike. First-hand explicit IC was already in the norm in major parts of the EU. Still, the pre-GDPR era includes subjective privacy-preserving healthcare metadata handling by healthcare providers. The term subjective implies that the healthcare providers' business goals subjectively dictated the data pipeline of the patients. Here, the loophole was that the business goal of healthcare organizations could have been starkly different from the objectives of the patients while receiving care services. Thus, the patients exercised limited powers concerning their healthcare metadata's purpose, usage, disclosure, and retention.

In post-GDPR era, the GDPR with its clause 7 disintegrates the integral aspects of IC. It defines IC as an explicit, autonomous, and automated decision-making and profiling at the onset of data collection. Moreover, IC gives the data subjects the right to access, object, erase, rectify and restrict their sensitive metadata sharing at any point in time. For metadata storage and sharing, the IC concerning issue is peer reviewed, readily received, and portable metadata (Article 20). The latter part promotes data utility and hampers the privacy-preserving measures. Privacy primarily has dichotomous aspects; first, the autonomous, explicit, and informed decision-making, and second is direct/indirect control over PII with powers to object, erase, rectify, and restrict the personal data processing at any point in time. Here, data utility satisfies three criteria: uninterrupted data collection, anonymization/pseudonymization, and uninterrupted data storage and sharing and is described in the context of IC in the following paragraph.

GDPR future expectations-based era: For an effective application of Informed Consent, a diagnostic lab (as biorepository and biodata bank) is selected as a case study for focused evaluation of the topic using content analysis. The idea of IC raises confusion because the patients are usually unaware of their healthcare metadata processing, primarily when some Trusted Third Party (TTP) researches the data, for example, AI machine vendors, clinical/scientific researchers, and the data analysts for IoT in healthcare. As disintegrated, anonymous processors ameliorate data transparency but lag in accountability. Accountability of Trusted Third Party (TTP) is vital for better health outcomes of the patients and the society at large. The EU suggests contract-based relationships with healthcare providers and TTP. The contracts must mention the description of TTPs' applied information security measures. In the contract, the signing bodies must mention the vital aspects such as the compensations at the data leak incident, inform about the data outsourcing, and delete the data after assigned data processing. In case of no contract signing, the healthcare provider would be held accountable. The regulatory and legislative authorities at the EU and national levels must unite to fill the TTP-based loopholes. For cryptography in healthcare, anonymization is one-way cryptography that is irreversible and removes personal identifiers and permanently deteriorates the performance of the analytical data. Pseudonymization is two-way cryptography where the identifiers are coded and are reversible, unlike anonymization [79]. Pseudonymization is a preferred privacy-preserving measure in healthcare, as it does not affect the healthcare data utility. The GDPR future expectations-based era requires the improved clause 7 with the implantation of a broad,

open-ended IC at the onset of data collection. Furthermore, for uninterrupted data collection, the issue is that the patients' IC is required in each phase with the exceptions of scientific studies. This exception applies, provided that the scientific studies are based on the public's vital and legitimate interest. To this end, the solution is to define public, vital, and legitimate interest and to ensure a broad, open-ended IC at the onset of the data collection.

3.2. At Inter- and Intra-Organizational Level

Results: For agent-centric, privacy-preserving bilateral relationships, the diagnostic lab (biobank and biorepository) and other key stakeholders are evaluated; see Figure 4.

To simplify understanding this rather complex healthcare network, the padlock chain model provides a linear and easily comprehensible model. This model represents the lab's bilateral relationship with other key-value actors in privacy-preserving healthcare data sharing. The types of value actors and value activities are similar in this model as well. All these actors perform their respective PbD measures. Similarly, privacy by policy is ensured: by international and national standard organizations and by domain-specific national regulatory authorities. Informed Consent (IC) is vital in initiating the privacy-preserving healthcare data sharing within and across healthcare providers. In the Netherlands, the majority of diagnostic labs are running their sister pharmacies. To share data with the pharmacies, the labs have to show, to the former, the proof of patients' IC. This proof is then countered with patients' PII shares. Here, IC is part of the PbD measures as well as of the privacy-by-policy measures seen in the model in Figure 4. The Dutch labs are obliged to take IC of the patients before collecting their biosamples. However, there is a significant loophole because the diagnostic labs' functioning initiates with an opt-out IC instead of an opt-in IC. This implies that patients can opt out of becoming data subjects for further scientific studies or clinical trials if they are explicitly expressed. Otherwise, they are considered as confirming the terms and conditions of becoming data subjects for clinical trials and clinical studies. Post IC, the lab's processing of patients' data is also subject to further scrutiny and rightly so [12,13].

Figure 4. Privacy assurance evaluation at inter-organizational level.

3.2.1. Evaluation

Informed Consent (IC) serves as a starting point for both PbD and privacy by policy. Therefore, it is an integral ingredient of PbD as well as privacy by policy measures in model 2 (see below) because it is obligatory upon lab (biobank and biorepository) to take informed consent at the onset of the patient's interaction with the lab or at the beginning of

the contract. The rest of the privacy-preserving procedures serve as succeeding steps to informed consent.

International and national standard organizations play a significant role in setting standard ISMS parameters for Dutch private labs in the Dutch healthcare setting. Non-governmental organizations (NGOs) such as the International Organization for Standardization (ISO) and NEN, Dutch Network of standardization, bind labs to follow privacy by policy rules and regulations. An international federation of national standards, ISO, facilitates standardization processes to support the international exchange of goods and services. NEN is the Dutch branch of ISO within the Netherlands for developing and applying standards nationally and internationally. These standard organizations and the national regulator prompt labs to instill privacy in their value network via policy into their system designs.

This e^3 -value model exhibits the value transactions from the organization's (i.e., diagnostic lab) perspective to maintain an equilibrium in safeguarding individuals' fundamental rights while sticking to their business goals. Business goals are the target business objectives of healthcare providers. The given models provide the basis for further research in first analyzing the standard parameters from nominal to ratio-based measurements to provide the standard metrics for key stakeholders to confirm the transfer of privacy as an objective value object within healthcare.

An important noticeable factor is the shift of data source from patients to organizations (healthcare providers). The source of data generation shifts from patients to the diagnostic lab when seen from the perspective of value objects and their data transfer. For example, when a patient goes to the diagnostic lab for a blood test, the results make the lab a source-generating point for related information/data. Then, the patient's power reduces in favor of lab management concerning the inflow and outflow of that information/data. Then, the privacy assurance measures such as IC, PbD, and privacy by policy are applied to further the interest of patients. Tackling that information largely depends upon the lab or, in other words, upon the privacy-protecting ontologies of the lab.

3.2.2. Patients' Un-Anonymized Metadata Share and Privacy Preservation: Conceptual Validation and REA Ontology

Sepsis Dataset Analysis using Process Mining and Expert's Opinion: The Sepsis event log is analyzed using PM tools (Disco and ProM Lite [63,64]) that validate the patients' un-anonymized metadata share amongst Dutch caregivers. Furthermore, an infield IT expert validated the veracity and applicability of PM experiments to real healthcare scenarios. IT expert is affiliated with a Dutch diagnostic lab and a local Dutch hospital. Afterward, the REA's Insurance Model (IM) [66] is extended. The model verifies who, how, and what leads to patients' un-anonymized metadata-share amongst the Dutch caregivers and explicates privacy as a materialized claim.

PM experiments findings and discussion: The commercial tool Disco (using fuzzy minor algorithm) discovered that there are 16 activities for 1050 cases and 15,214 events/instances. In Figure 5, the rectangles are the activities or the tests on Sepsis patients by different suborganizations within the hospitals. The black arrows show the patients' metadata-share amongst the horizontally distributed suborganizations within the hospital with respective time stamps. A higher number of patients show thicker arrows and vice versa. The number of patients and respective data sharing time are demonstrated along each arrow. Furthermore, the process model with sharing time stamps is given; see Figure 5. In the given legend, the subhospital departments with pseudo-identifiers are shown with activities share in percentage. The periods of care data sharing amongst horizontally distributed caregivers and their frequency are noted (with activities' median and least time stamps). The experiments suggest that the succeeding data sharing is performed either instantly or after a short period from one suborganization to the other. PM experiments explained that no privacy-preserving actions could practically have occurred in such brief time frames.

(a)

(b)

Figure 5. (a) Sepsis process model (using Disco tool): patients’ un-anonymized care metadata share amongst horizontally distributed caregivers; **(b)** Legend for Sepsis process model.

ProM Lite: In Figure 6, the dotted chart algorithm from the ProM Lite tool is used on the Sepsis event log. The dotted chart depicts the undergoing of activities with a further detailed description of time stamps. On the x-axis, we see the time duration of the Sepsis tests and data share amongst horizontally distributed organizations within the hospital. Whereas on the y axis, we see the trace of events of each patient. A similar color depicts a similar Sepsis test (activity). In the legend, each color is manifested with a respective activity in the graph; see Figure 6. The PM experiments suggested that the diagnosis consumes the maximum time for patients' data to move from one suborganizational department to another. Whereas the majority of times, the transfer of patients' data was performed either instantly or within 48 hours, as in Figure 6. This explained that no privacy-preserving measures could have occurred in this short while.

Furthermore, in ProM Lite, the social network algorithm discovers information regarding working (medical/administrative) staff in an event log. We could not reveal the social network in the Sepsis event log using ProM. The absence of a social network suggested that either the staff disagreed with sharing their PII publicly or it was intentionally withheld for privacy's sake. However, the Dotted Chart substantiated our first finding of Disco. Later, the IT expert also verified that while sharing patients' PII with vertically distributed organizations, healthcare providers restrict the information share of their employees for privacy's sake.

Expert Opinion: An IT expert working in a local Dutch hospital was shown the PM experiments and results and was asked whether caregivers share un-anonymized (and pseudonymized) care data amongst horizontally distributed caregivers. Furthermore, do they remove the PII of medical and administrative staff? We also asked to confirm whether the results are generalizable to the vertically distributed Dutch caregivers or not. The IT expert validated all our findings.

Significantly, the REA ontology verified the underlying economic factors behind patients' un-anonymized metadata share amongst Dutch caregivers, of better data utility.

Figure 6. Privacy assurance evaluation at intra-organizational level (Sepsis event log using the Dotted Chart algorithm from the ProM Lite).

3.2.3. Conceptual Modeling Framework Using REA Ontology

Results: Dutch caregivers are privately run and partially publicly funded enterprises. Our goal behind using REA ontology was to discover the underlying financial priorities of care enterprises for patients' un-anonymized metadata share [66]. REA's Insurance Model (IM) is an extended application model because it includes contract and commitment levels to the increment and decrement events between economic agents for mutual value gain or loss. Before conceptualization fundamentals, it is vital to emphasize that the PM experiments in Figures 5 and 6 verified our assumptions regarding patients' un-anonymized metadata share amongst Dutch caregivers. Whereas Figure 7 is about the proposed REA model, which relates to real-life events where privacy is used as a Materialized Claim against caregivers/health insurers in the Netherlands. Figure 7 is explained from top to bottom, describing the concepts and relations of the model below:

Figure 7. REA using Insurance Model and explicating privacy as a materialized claim.

The economic agents' are legal entities who lose or gain control over the economic resources through economic events/interactions. An economic resource is a thing or service to be planned, monitored, and controlled by the concerned authorities (economic agents). Here, resources such as cash, metadata, and care services are exchanged as increment or decrement events. The increment is the inflow of resources, while the decrement is the outflow of resources from a patient's perspective. Contracts are legal commitments that involve each agent as a party. Initially, the contract is signed between two active

‘party’ agents. The passive ‘party’ agents activate their value transactions later with future increments/decrements events. For example, the passive agents include the caregivers who are absent at signing the insurance contract but later become active at their service delivery. Initially, the patients and health insurers perform increment and decrement events by exchanging cash with and for one another. With an insurance contract, the PII of the insured is also stored in insurer IS. This information intake (economic resource) leaves the insured with less control over his or her PII (for physical, decisional, and informational security). The insurance contract clauses commit the (party) agents for future increment and decrement events (including the privacy preservation of the insured).

From the patient’s perspective, the increments include the timely cash payments from the insurer to caregivers, clinical care to the patient, and the patient’s PII security assurance from insurer and caregivers alike. The decrement events are monthly insurance payments from the patient to the insurer and the outflow of patients’ PII (as an input to metadata) to the insurer and caregivers. After the increment and decrement events’ execution, the involved agents evaluate any imbalances between the materialization and the settlements of the patient’s resources. Patients usually evaluate their physical security/recovery in comparison to their cash payments to the insurer. Materialized Privacy Claim (MPC) surfaces only when either the patient or any other involved potent authority (such as Data Protection Officer) claims to assure the patient’s informational and decisional security in addition to his/her physical security/recovery. For instance, recently, the Dutch Data Protection Officer (DPO) fined Haga hospital 460,000 euros for a Dutch celebrity’s privacy breach [80] and charged Menzis (health insurer) euros 50K for care data mishandling [81].

Evaluation: Privacy and metadata in REA Ontology: In REA ontology, the metadata, the formation, maintenance, and interoperability is part of an organization’s ‘posting and dimension aspect of financial disbursement’ instead of its application model [66]. Earlier (in 2006), metadata handling was considered dependent upon the behavioral pattern of the respective organization [66]. Although the metadata entries are always stored using the identity strings (ID strings with PII) in organizations’ ISs. Still, they lacked standardized information security management systems [35]. Nowadays, metadata privacy is protected by national and European regulations and involves legal commitments by the concerned authorities to avoid hefty money claims [6,7]. The legal and administrative requirements for privacy include privacy by policy, privacy by design, and patients’ informed consent measures [6,7]. Therefore, the privacy preservation of care metadata share does not rely on the behavioral patterns of health insurers/caregivers anymore. Rather, legal benchmarks constrain the concerned authorities for shaping the respective organizational/services contracts accordingly. Consequently, the patients’ privacy concerns now appear as Materialized Privacy Claims (MPCs) in the Dutch care metadata share landscape.

Another approach (for a better generalizable REA model) was to incorporate privacy breach as a condition in the insurance policy. Furthermore, it could have added granularity regarding the levels of severity and respective cash disbursements. Nevertheless, the proposed REA model signifies the scenarios involving privacy claims by patients (or for patients by any other potent authority, i.e., DPO) against caregivers and insurers. In this regard, the possibility of these privacy claims being compensated with hefty cash payments by the concerned authorities (caregivers/insurers) is high; see Figure 7. The red dotted blocks show our contribution to the model.

Financial Gain/Loss of Healthcare Enterprises and Healthcare Metadata Utility: If you see Figure 8, the healthcare providers such as insurers, GPs, hospitals, and diagnostic labs along with their sister pharmacies are financial enterprises. Enterprises that are privately run but are partially publicly funded and are sometimes nonprofitable.

We discover that financial gain/loss depends directly upon the healthcare data utility. Data utility is also vital for patients’ efficient and effective treatment and data spaces integration across domains in healthcare both nationally and internationally for collective gain. We are also working in this regard to find out both technical and managerial

measures to locate optimal data utility while preserving patients' privacy while sharing patients' metadata.

Figure 8. Financial gain/loss of healthcare enterprises and healthcare metadata utility.

3.3. At Third-Party Data Processing Level

For third-party data processing, we focused on the PM domain using real-world healthcare event logs.

Third Party Data Processing in PM domain using publicly available PM tools and techniques on real-world Healthcare Datasets: This research verifies the level of privacy preservation of noise-adding plugins in the publicly available Process Mining tool ProM. The privacy preservation is evaluated along with their respective data utility. The data utility of the plugins is verified using quality indicators' of the deduced (noise-added) business process models called Petri Nets. Furthermore, the noise-added Petri Nets are compared against the original Petri Nets for assessing the increase/decrease: in both the privacy and data utility of the event logs.

Process Mining and Data Analytical Approach: For an overview, an illustration of our data-analytics approach for third party privacy assurance evaluation is given in Figure 9. The original event logs were made more privacy preserving [82] by using 12 noise-adding plugins in two versions of ProM (6.9–6.10). For both the event logs and in both the ProM versions, two similar (plugins producing TM1 and TM6) and ten different (five plugins each) noise-adding plugins were evaluated based on respective quality indicators (explained below) of the noise-added Petri Nets, i.e., Training Models (TMs). For each test model of the original event log, the inductive miner plugin was used to deduce the test model's Petri Net in both versions of the ProM tool. An inductive miner was specially used to maintain the soundness of the event logs [83], also because the inductive miner nicely evaluated the event logs and gave good performing Petri Nets with higher utility indicators' values, i.e., fitness, precision, simplicity/complexity, and generalization [83]. Conformance checking (utility) indicators, i.e., fitness, precision, and complexity, were

evaluated using ProM with ‘replay, conformance checking, and metrics analysis’ plugins. This allowed us to skip manual (visual) evaluations.

Figure 9. Data analytical approach for third party privacy assurance evaluation.

Fitness (domain: $0 \dots 1$) quantifies the level of observed behavior in the derived model compared to the event-log from 0 to 1. Good fitness indicates a convenient replay of the behavior seen in the event log [47,48,53,84]. Fitness is computed based on all three fitness matrices: move-model fitness, move-log fitness, and move-trace fitness. Precision (domain: $0 \dots 1$) computes that segment of the model’s behavior that is not visible in the event-log from 0–1. Precision is computed based on all three precision matrices, i.e., overall precision along with backward precision and balanced precision [84]. Complexity itself is an inverse of the quality indicator, simplicity. Simplicity is quantified by its inverse, i.e., complexity. Complexity is measured using the Extended Cyclomatic Metric, i.e., ECyM, and Extended Cardoso Metric, i.e., ECaM of Petri Nets, independent of the event log. The higher the value of these matrices, the lower the simplicity level of the respective model. A higher complexity level implies more errors [53].

The step-wise procedure to quantify the quality indicators for the real-world test models is as given below:

For fitness: In ProM, the Sepsis/hospital billing event-log is imported → deduce the test model, using, Mine a Petri Net using Inductive Miner → the Petri Net is replayed along the original Sepsis/hospital billing event-log using Replay a log on Petri net for conformance checking → concept name, and ILP based replayer assuming 32,767 tokens are opted → for the rest, the default setting is kept → in visualization, the PNET alignments setting → in global statistics the move-model fitness, move-log fitness, and trace-fitness are measured → visualizing, project alignment to log, gave log’s each trace alignment with the model.

For precision: Conformance checking is performed with the plugin check precision with ETConformance, → ordered representation → for the algorithm, align precision using ILP-based replayer assuming at most 32,767 tokens → afterward, and the default setting for set parameters is selected → quantified the overall precision, backward precision, and balanced precision.

For Complexity: Two matrices, namely ECyM and ECaM, are measured. Both of these are measured using the plugin Show Petri Net Metrics in ProM. The research expectation is: in evaluating PUT of noise-adding plugins in ProM, if a training model shows similar or higher fitness and precision with changed complexity values, then the training model shows higher data utility. Crucial factors to remember while evaluating are as follows:

- Proximity to higher fitness values (toward fitness value 1) is essential for higher data utility evaluation, whereas, for a second indicator, any remaining quality indicators add up to the process evaluation.
- Only precision and fitness automatically evaluate the suitability of the deduced Petri Net in comparison to the original event-log in ProM.

- Complexity is still measured because it quantifies Petri Nets' metrics irrespective of the event log. For example, suppose the values of the Extended Cyclomatic Metric (ECyM) and Extended Cardoso Metric (ECaM) are similar to the test models. In that case, it suggests that the noise-adding plugins are showing unaffected models (without noise). Unaffectedness is further confirmed if precision and fitness values of TMs also coincide with that of the test model. For this reason, the visual comparison between the models is substituted by calculating the respective complexity of the deduced models, i.e., the test models and TMs.

Experiments Results and Evaluation

This section comprises ProM experiments, expectations evaluation, and result discussion.

ProM Experiments Results: The quality indicators: fitness, precision, and complexity were quantified for the 2 test models and 12 Training Models (TMs). Two distinctive subgroups of noise-adding plugins and their respective training models were attained. One subgroup showed similar or higher fitness and precision values and changed complexity values from that of the test models named as 'TMs with high data utility' and 'TMs with low data utility' subgroups showed lower or similar indicators' values. The blue-colored TMs with high data utility subgroup comprises the TMs 2, 4, and 6 in Sepsis event log using ProM 6.9 and TMs: 6, 8, 9, 10, and 11 for hospital billing event log using ProM 6.10. For a detailed description, see Tables 1–3.

Table 1. ProM experiments results for the 2 test models and 12 training models using noise-adding plugins (Fitness).

Quality Indicators		Fitness		
Respective Metric Names		Move-Model	Move-Log	Trace
Test-Model (Sepsis)	Raw Event-log Model	0.9265	0.9253	0.9030
Training Models (obtained using ProM 6.9 noise-adding plugins)	TM ₁ (20 % noise)	0.9842	0.9249	0.9249
	TM ₂ (add chaotic-events)	0.9264	0.9224	0.9009
	TM ₃ (add frequent chaotic-events)	0.9265	0.9253	0.9030
	TM ₄ (add infrequent chaotic-events)	0.9265	0.9253	0.9030
	TM ₅ (swap trace-events)	0.9636	0.9458	0.9325
	TM ₆ (low frequency filtered-out)	0.9773	0.7271	0.7756
	TM ₇ (high frequency filtered-out)	0.9265	0.9253	0.9030
Test-Model (Hospital billing)	Raw Event-log Model	0.9908	0.9084	0.9218
Training Models (obtained using ProM 6.10 noise-adding plugins)	TM ₁ (20 % noise)	0.9994	0.9707	0.9742
	TM ₈ (Anonymized Event-log)	0.9908	0.9083	0.9217
	TM ₉ (Obfuscate lit-event-attrib-values)	0.9908	0.9083	0.9217
	TM ₁₀ (Obfuscate event-attrib-name)	0.9915	0.9067	0.9218
	TM ₁₁ (Add/change concept)	0.9908	0.9084	0.9218
	TM ₆ (low frequency filtered-out)	0.9954	0.9409	0.9481
	TM ₁₂ (reversed log)	0.9902	0.4034	0.4947

Table 2. ProM experiments results for the 2 test models and 12 training models using noise-adding plugins (Precision).

Quality Indicators		Precision		
Respective Metric Names		Precision	Backward	Balanced
Test-Model (Sepsis)	Raw Event-log Model	0.5353	0.3989	0.4671
Training Models (obtained using ProM 6.9 noise-adding plugins)	TM ₁ (20 % noise)	0.3864	0.3135	0.3499
	TM ₂ (add chaotic-events)	0.5008	0.3868	0.4438
	TM ₃ (add frequent chaotic-events)	0.4125	0.337	0.3747
	TM ₄ (add infrequent chaotic-events)	0.513	0.3956	0.4543
	TM ₅ (swap trace-events)	0.3288	0.2798	0.3043
	TM ₆ (low frequency filtered-out)	0.7924	0.7253	0.7589
	TM ₇ (high frequency filtered-out)	0.5346	0.398	0.4663
Test-Model (Hospital billing)	Raw Event-log Model	0.6691	0.5844	0.6268
Training Models (obtained using ProM 6.10 noise-adding plugins)	TM ₁ (20 % noise)	0.3526	0.2699	0.3112
	TM ₈ (Anonymized Event-log)	0.787	0.6576	0.7223
	TM ₉ (Obfuscate lit-event-attrib-values)	0.7232	0.666	0.6946
	TM ₁₀ (Obfuscate event-attrib-name)	0.6776	0.6309	0.6543
	TM ₁₁ (Add/change concept)	0.7709	0.6256	0.6983
	TM ₆ (low frequency filtered-out)	0.9954	1.0	0.9841
	TM ₁₂ (reversed log)	0.7543	0.8782	0.8163

Twelve noise-adding plugins along with their respective names, type of noise addition, and the capability of better utility/privacy preservation are given in Table 4. Usually, the term ‘further anonymization’ is used when pseudonymized healthcare datasets are made further privacy preserving by adding noise. Similarly, in this study, the raw Sepsis and Hospital billing event logs were anonymized using noise-adding plugins in ProM 6.9 and ProM 6.10, respectively. The purpose of using different event logs in two different ProM versions was to evaluate the noise of adding plugins in different ProM versions using different healthcare event logs. To verify the findings, two similar plugins were also evaluated in different ProM versions with different event logs where both plugins (i.e., 1 and 6; see Tables 1–3) were judged based on the maintenance of their respective utility levels. The aim is to measure the difference between the data-utility of further anonymized training models compared to that of the original test model in two different versions of ProM.

Table 3. ProM experiments results for the 2 test models and 12 training models using noise-adding plugins (Complexity).

Quality Indicators		Complexity	
Respective Metric Names		ECyM	ECaM
Test-Model (Sepsis)	Raw Event-log Model	295	24
Training Models (obtained using ProM 6.9 noise-adding plugins)	TM ₁ (20 % noise)	825	24
	TM ₂ (add chaotic-events)	319	25
	TM ₃ (add frequent chaotic-events)	714	28
	TM ₄ (add infrequent chaotic-events)	307	25
	TM ₅ (swap trace-events)	3472	28
	TM ₆ (low frequency filtered-out)	204	21
	TM ₇ (high frequency filtered-out)	295	24
Test-Model (Hospital billing)	Raw Event-log Model	1946.0	35.0
Training Models (obtained using ProM 6.10 noise-adding plugins)	TM ₁ (20 % noise)	10011.0	37.0
	TM ₈ (Anonymized Event-log)	897	34
	TM ₉ (Obfuscate lit-event-attrib-values)	897	34
	TM ₁₀ (Obfuscate event-attrib-name)	984.0	34.0
	TM ₁₁ (Add/change concept)	454	33
	TM ₆ (low frequency filtered-out)	16	12
	TM ₁₂ (reversed log)	47	25

Table 4. 12 noise-adding plugins, their noise addition types, their and utility vis-a-vis privacy levels.

Noise-Adding Plugin	Plugin's Name	Type of Noise Addition	Higher Utility	Higher Privacy
1st	Add noise to log filter	20% events' removal & addition	×	✓
2nd	Add chaotic events to log	chaotic events' addition	✓	×
3rd	Add chaotic activity to log; frequency-dependent	Addition of k frequent randomly positioned activities	×	✓
4th	Add chaotic activity to log, frequency-independent	Addition of k uniform randomly positioned activities	✓	×

Table 4. Cont.

Noise-Adding Plugin	Plugin's Name	Type of Noise Addition	Higher Utility	Higher Privacy
5th	Add chaotic activity to log, infrequent	Addition of k infrequent randomly positioned activities	×	✓
6th	Low-frequency-filtered log	Outliers removal	✓	×
7th	High-frequency-filtered log	Filtering out high-frequency activities	×	✓
8th	Anonymized event log	Obscuring of event attributes	✓	×
9th	Obfuscate literal event attribute values	Obscuring of event attributes (trace changing)	✓	×
10th	Obfuscate event attribute names timestamps	Obscuring of event attributes' names (trace changing)	✓	×
11th	Add/change concept; name of event from classifier	Addition/changing of events's names	✓	×
12th	Reversed log	Reversing of trace	×	✓

3.4. Twelve Noise-Adding Plugins and Their Privacy Utility Tradeoff

Expectations Evaluation including Discussion: The ‘TMs with high data utility’ results verify the aforementioned underlying reasons (see research expectation in Section 3.3). The blue-colored TMs (see Tables 1–3) give similar or better fitness and precision values and different complexity values to that of the test model. The TMs with lower data utility show either lower fitness values (the essential first step in data quality assurance) or lower precision values [84] or similar fitness, precision, and unaffected complexity. Lower complexity shows the higher simplicity and simplified Petri Net, but it is less of a concern here. However, the differing complexity values are important in diagnosing that the deduced process models added noise and were not identical to the test models. The blue-colored TMs results in Tables 1–3 verify our research expectation, “evaluating PUT of noise-adding plugins in ProM. The **research expectation** is: in evaluating PUT of noise-adding plugins in ProM, if a training model shows similar or higher fitness and precision with changed complexity values, then the training model shows higher data utility.”. However, does this evaluation imply that the respective plugins will always give higher data utility for each event log? The performance of plugins highly depends upon the quality and domain of the event log. An event log with (comparatively) less sensitive and nicely put data would allow the plugin to give better data utility (Hospital billing is comparatively lesser sensitive and gives better data utility see TM1 and TM6). This was further confirmed by evaluating another publicly available (healthcare) event log (i.e., Urineweginfectie (UWI-casus) logboek) [65]. The event log gave around 0.1 precision and 0.6 fitness for its test model using similar plugins with Prom 6.10. Still, the TMs with higher data utility would give similar or higher fitness and precision values (i.e., 0.1 precision and

0.6 fitness) and vice versa. Thus, the respective utility indicators' values vary but not the data utility of the plugins.

Threats to Validity: Plugins' performance depends directly upon the quality and domain of the event log. Plugins give better data utility to the more suitable event logs, especially when trained algorithms within the plugins nicely find integrated information from the event log and efficiently perform their tasks. Domains with less sensitive event logs allow the plugins to give even better data utility. Still, the plugins with higher data utility maintain better data utility on the less suitable (lower quality) event log and vice versa. The conclusion of the research work relies upon the correct usage of the mentioned plugins. Using different plugins (to generate Petri Nets) might suffice if similar Petri Net generating plugins (here, i.e., inductive miner) are used for both (the test model and the TMs).

ProM is the one-stop, publicly offered tool with a comprehensive range of plugins. It is worthwhile to explore its offerings from the perspective of the end-users. Multiple other PM tools are commercially available, but so far, ProM is generally relied upon across the board (especially by academia). The experiments are conducted on different event logs and two versions of ProM (i.e., ProM 6.9 and 6.10). This research work aims to facilitate PM end-users in availing privacy-preserving plugins with high data utility. This is also performed in the context of suggesting similar noise-adding plugins for future ProM versions.

So far, in PM, the PUT evaluation has been considered in the context of case attributes. Nevertheless, other confidentiality risks, such as trace disclosure, had not been given much attention. As single-trace identification is sufficient for an adversary to locate an individual from the event-log [36,44]. In this regard, plugins are incorporated in ProM 6.10 for further anonymization of the event log that gives higher data utility; see plugins 9, 10, and 12 in Table 4.

4. Main Conclusions and Contribution: Multilevel Privacy Assurance Evaluation and Privacy Utility Tradeoff

Figure 10 depicts the multilevel privacy assurance evaluation. The research work also evaluates the impact of the privacy utility tradeoff on multilevel privacy-preserving measures. The multilevel privacy-preserving measures include individual, inter-/intra-organizational, and third-party metadata processing levels see.

Figure 10. Contribution: Multilevel Privacy Assurance Evaluation and Privacy Utility Tradeoff.

Main Conclusion: The normative and empirical evaluation is elaborated for each level. The content analysis using the literature review, official websites, and online official documents was performed for normative evaluation. For empirical evaluation, Process Mining was used on real-world healthcare event logs. Normative and empirical evaluations are then illustrated using either e^3 value modeling or REA ontology for individual and inter-/intra-organizational levels. The e^3 value modeling and REA ontology were then further validated with expert opinion. PM is performed using real-world healthcare datasets for the third-party data processing level to explore the privacy-preserving data utility-prone PM plugins. Privacy preservation is attained by adding noise using PM noise-adding plugins. At each level, the focus is drawn upon the privacy utility tradeoff and its influence upon the privacy increase/decrease. Based on content analysis, the padlock chain model of e^3 -value modeling is drawn at the individual level. Generally, to draw an e^3 -value model, usually physical or tangible value transactions are demonstrated and are in contrast to the qualitative value transactions such as privacy, security, transparency, autonomy, and confidentiality. As the latter are all respective terms and depend primarily upon the purposes and requirements of the involved stakeholders. The e^3 -value model verifies privacy as transforming from a subjective value to a (subjective) value object in a value model. It satisfies all four attributes to become a value object in the padlock chain model. Privacy is: a business goal of healthcare enterprises, is accredited by global and national standard authorities, is exchanged against PII of the patient, and is regulated by national regulatory authorities. Patients' Informed Consent is crucial in initiating the collection and processing of Patients' PII. IC is required in each phase, with exceptions to scientific studies based on vital and legitimate public interest. The definition of the vital and legitimate public interest is required to avoid confusion. Moreover, a broad, open-ended IC is required for healthcare providers to extract BDA's value further. BDA and data utility are vital for efficient healthcare services and better healthcare outcomes. With normative evaluation and e^3 -value modeling, privacy is visible as a (subjective) value object.

However, privacy has not yet turned into an objective value object because the quantitative, standard evaluation of privacy by design and privacy by policy measures is still missing in fear of losing data utility (i.e., Privacy Utility tradeoff). One such example occurred, in January 2021, when the data theft was reported by local health group GGD from its CoronIT system. The CoronIT system contained the sensitive data of tens of thousands of Netherlands residents who underwent a coronavirus test. The criminals gained access to the data by bribing GGD call center employees and making sensitive data available for months for online trading. The data comprised the patients' addresses, telephone numbers, email addresses, and citizen service numbers, i.e., BSN). Ironically, a month earlier, the Dutch health minister publicly appreciated the CoronIT system for being easily accessible with its export function, where the downloading and forwarding of personal data was deliberately made as a built-in function for easy report making and sharing with other systems. In another statement, the similar CoronIT system was censured for its vulnerability against privacy risks, where he demanded to maintain better identity and access management to help prevent unauthorized users from gaining access to specific sensitive data [85]. This example perfectly manifests the redundancy of Informed Consent when FAIR (Findable, Accessible, Interoperable, and Reusable) becomes the need of the hour for an effective and efficient healthcare during infectious diseases such as COVID-19. Moreover, it is also evident that privacy measures are often neglected for better data utility in handling sensitive healthcare metadata for better healthcare outcomes.

At inter-/intra-organizational level, for better insight, one of the key healthcare providers is analyzed to evaluate the privacy-protecting bilateral relationships between lab (biobank and biorepository) and other key stakeholders using padlock chain model. In e^3 -value models, the actors are divisible as principal, agent, third party, and regulator. All these actors do value transactions as front-end, counter, back-end, and regulation, respectively. The actors utilize their respective "privacy by design" parameters instilled in their respective Information Systems. Privacy by policy is evaluated by international and

national standard organizations along with domain-specific national regulatory authorities. Informed consent plays a vital role in completing the privacy-preserving healthcare data sharing within and across healthcare providers. Labs are obliged to take IC of the patients before collecting their biosamples. The IC in labs is based on an opt-out procedure instead of an opt-in procedure. The patients can opt-out of becoming data subjects for further scientific studies if they explicitly say so. Otherwise, they conform to the terms and conditions of becoming data subjects for clinical studies. Once granted informed consent, the lab's processing of patients data is also subjected to further scrutiny and justly so.

A real-world Sepsis dataset was analyzed with Process Mining (PM) discovery techniques using Disco and ProM Lite to validate the un-anonymized clinical care metadata-share amongst Dutch caregivers. The experiments' results verified that the horizontally distributed Dutch caregivers share un-anonymized patients' metadata. An IT expert (affiliated with a local Dutch hospital and a Dutch diagnostic lab) further verified the PM results and their generalizability to the vertically distributed Dutch caregivers. Furthermore, the PM results and the IT expert also confirmed that the PII of the administrative and medical staff is intentionally removed during the metadata share across caregivers from different domains for privacy's sake. In this regard, the REA model helped us discover who, how, and what leads to an un-anonymized metadata-share amongst vertically and horizontally distributed Dutch caregivers.

REA's Insurance Model (IM) uncovered the key economic agents, their prime economic interactions in the Dutch care metadata share landscape from the patient's perspective. The health insurer and patient are key (active) economic (party) agents who sign the contract of health insurance where caregivers are initially the passive (party) agents. The (party) agents become committed to the contract clauses for the increment (inflow) and decrement (outflow) of economic resources. As an increment, the patient receives the cash inflow from the insurer to the caregivers (who activate with respective contracts registration by the patient) on each medical checkup visit, the care services from caregivers and preservation of PII (via contracts). In decrement, the patient pays monthly payments to the insurer and gives his or her PII as an input to the insurer and caregivers' metadata. The decremented resources become valuable to the receiver. The receiver then enjoys the authority to evaluate the imbalance (if there is any) between the materialization of economic resources. The economic resources are provided care for the physical, informational, and decisional security and their prior settlements such as cash payments, contract clauses ensuring privacy). A significant result of our extended REA IM model is that the privacy concerns take the form of a Materialized Privacy Claim (MPC). MPC emanates when an economic agent (either the patient or any other potent authority) finds the imbalance in patients' exchanged resources. This modeling approach explains the who, how, and why of money claims for illegal information disclosures as MPC. Recent MPCs include 460,000 euro against the HAGA hospital, 440,000 euro against OLVG Amsterdam hospital [80,81,86], and 50,000 euros against Menzis insurance company by Dutch DPO. These incidents also suggest that data-utility-prone healthcare metadata handling is in practice in the Dutch healthcare landscape for better healthcare outcomes.

At the third-party data processing level, the empirical evaluation is performed using PM tools and techniques on real-world healthcare event logs. In Process Mining, three basic information pieces are crucial for revealing important process information concerning the event log (dataset). For that, the event should be primarily connected with a case (i.e., individual), an activity, and two-time stamps (initiating and ending timestamps). A sequence of events for one case/individual is called a trace. Identifying one trace is enough for an adversary to exploit the PII of a data subject (individual). This research work utilized noise addition for privacy preservation of the data subjects to the real world event logs and checked its influence upon the data utility of the event logs using ProM.

The noise addition for privacy preservation to the level of maximum data utility is well acknowledged in the field of differential privacy. The process requires time, money, and effort in the long run as it requires the tools to be inherently differentially private. This

research provides a short-term, end-user-friendly approach where existing PM tools and noise-adding techniques are used for noise addition. The experiments revealed some noise-adding plugins that provide good data utility as well. The performed PM experiments add noise by interchanging the sequence of traces, cases, or activities. The detailed description is provided in Section 3.4 and in Table 4. The resulting event logs are impossible for anyone to locate the right person for any activity. The solution is short-term but worth performing for PM end-users who are interested in doing privacy-preserving healthcare data analytics.

For data utility, the conformance checking indicators are evaluated using noise-adding plugins in the ProM tool. Thus, the indicators; fitness, precision, and complexity are evaluated by noise-added Petri Nets and respective noise-adding plugins using the ProM tool. The step-by-step procedures to evaluate fitness, precision, and complexity are given in Section 3.3. For fitness: move-log, move-model, and trace fitness, for precision: precision, backward precision, balanced precision, and for complexity: ECyM and ECaM matrices are evaluated, respectively. Fitness and precision values are measured using ProM because they calculate the comparative values between the deduced Petri Nets and the original event logs. Moreover, the varied values of complexity matrices exhibit the noise addition to the originally deduced test Petri Net. Using ProM, we found that the Petri Nets with higher data utility have similar or higher fitness and precision values with changed complexity values. Based on our findings, we further conclude that well-structured event logs give higher utility values. Nevertheless, this does not hamper or affect the performance of noise-adding plugins.

We evaluated 12 noise-adding plugins of ProM using real-world Sepsis and Hospital billing datasets. The ProM experiments gave us two subgroups of plugins based on similarity/dissimilarity between respective Training Models (TMs)—namely, TMs with high data utility displaying somewhat similar or higher fitness and precision values to the original test model but changed complexity values. The TMs with low data utility exhibited lower fitness and precision values than the respective test models or with similar (complexity) indicators' values. Plugins add chaotic events to log and add an infrequent chaotic event to log, and low frequency filtered out showed high data utility compared to the test model of Sepsis event log using ProM 6.9. For the Hospital billing event log, ProM 6.10 revealed that the: anonymized event log, obfuscate literal event attribute values, obfuscate event attribute name, add change concept, and low frequency filtered out gave similar or higher data utility. The findings are based on similar or higher fitness and precision values and different complexity values from that of the respective test model. TM₆ was produced using: low-frequency filtered-out plugin and was used for both the event logs and gave maximum data utility in both the logs with ProM 6.9 and 6.10, respectively. Moreover, the TM₁, add noise to log filter, was also produced with a similar plugin that was evaluated on both different logs using two different versions of ProM (i.e., 6.9 and 6.10) and gave a similar result of lower data utility than respective test models. The black-colored noise-adding, privacy-preserving plugins badly affected data utility with increased privacy preservation.

Table 5 provides a concise description of the empirical and normative evaluation of all three levels of multilevel privacy assurance.

The research provides the multilevel privacy assurance evaluation of fundamental privacy-preserving measures for sensitive healthcare metadata handling in the Dutch healthcare landscape. Table 5 provides the crux of the research work with the usage of respective techniques for each level ranging from individual to inter-/intra-organizational to the third-party data processor level. As Table 5 shows that for the first level, we provided the normative evaluation and its illustration with e^3 -value modeling, which was later verified with expert opinion. At this level, the empirical evaluation was not performed because the Informed Consent in its essence (for basic aspects see Figure 11) deteriorates with each level from individual to inter-/intra-organizational to third-party data processing level.

Table 5. Multilevel privacy assurance evaluation: using empirical and normative evaluation of fundamental privacy preserving measures for healthcare metadata handling.

MultiLevel Privacy Assurance Evaluation	Normative Evaluation (Content Analysis)	Empirical Evaluation (Data-Driven Process Mining)	e ³ Value Modeling	REA Ontology	Expert Opinion	Privacy Increase/Decrease (PUT Impact)
1st: Individual Level	✓	×	✓	×	✓	privacy increase (highest level)
2nd: Inter-/Intra-Organizational Level	✓	✓	✓	✓	✓	decrease
3rd: Third-Party Data Processing Level	✓	✓	×	×	×	further decrease

Figure 11. Multilevel Privacy Assurance Curve.

For the second inter-/intra-organizational level, we provided both normative and empirical evaluation and illustrated the findings using e³-value modeling and REA ontology, respectively. Both evaluations were verified by (in-field) expert opinion. For the third level of third-party data processing, the empirical evaluation was performed using PM with real-world healthcare event logs. Furthermore, normative evaluation is performed using the study of the literature, official websites, and online material. Our research work verifies

the significance of PUT in shaping the privacy status quo for healthcare metadata handling in the Dutch healthcare landscape.

4.1. Multilevel Privacy Assurance Curve

The below-given curve is a hypothetical multilevel privacy assurance path partially inspired by the aforementioned normative and empirical evaluations. The Figure 10 shows the multilevel privacy assurance path across two axes. The x-axis shows the level of individual control over one's PII in the Dutch healthcare landscape. The y-axis shows the level of inter- and intra-organizational control over data subjects' PII. The figure suggests that the individual control over Personally Identifiable Information (PII) is maximum before sharing sensitive personal information with a healthcare provider at the onset of data collection. Once the informed consent is taken from the healthcare provider, the control over personal information starts reducing. This control becomes weaker when, at first, the information is shared amongst horizontally located (intraorganizational) and then with vertically distributed interorganizational caregivers. The least individual control is when the patient's PII is used for third-party metadata processing for further data evaluation. Each quadrant suggests a stage of patients' value care where stage 1 (S1) is when she or he gives informed consent at the onset of data collection. Stage 2 (S2) is when healthcare metadata are shared across horizontally distributed caregivers, and patients' treatments are started. Stage 3 (S3) is when patients' PII is shared across vertically distributed caregivers, and the patient is treated, and the case is closed. The time for the latter situation can vary as per each patient's condition. Stage 4 (S4) is when a third-party data processor shares the stored, interoperable healthcare metadata for the sake of further data evaluations. At this stage, the individual's control over PII is the least and that of the organization (i.e., healthcare provider). Because then the Trusted Third Party (TTP) becomes potent and can exploit the available information to his or her end. The dotted area at the peak is when DPO can launch a Materialized Privacy Claim against healthcare providers to compensate for privacy breaches; see Figure 11.

4.2. Contribution

Our research work **contributes** in to the following areas:

1. Individual level: provides clarity regarding privacy definition and its qualitative assessment measures. Policymakers need to standardize the Privacy by Design measures, and entrepreneurs have to apply standardized technical measures to safeguard privacy to avoid hefty compensations effectively. The e^3 -value models provide the basis to provide the standard metrics for key stakeholders to confirm the transfer of privacy as an objective value object within healthcare. We highlighted the Informed Consent and its loopholes in ascertaining patients' privacy, as in Figure 10;
2. Inter-/intra-organizational level: validated the patients' un-anonymized care metadata shared amongst Dutch caregivers using Process Mining (PM) discovery techniques on a real-world Sepsis dataset. From a patient's perspective, we provide the REA ontology based on the content analysis concerning Dutch care metadata share landscape. IM of REA ontology highlights: the key economic agents, their prime interactions, and collective economic value gain or loss. We explicated privacy as a Materialized Privacy Claim (MPC) using REA;
3. Third-party metadata processing level: validated the usefulness of the plugins as privacy-preserving noise adding plugins for the tool's end-users from both the domains (i.e., the PM and healthcare). The contribution of this research work is to suggest and verify a framework for the evaluation of the PUT of ProM's (noise adding) plugins and their respective noise added Petri Nets. This was also conducted in the context of suggesting similar noise-adding plugins for future ProM versions;
4. Normative and empirical evaluation: provides a normative and empirical evaluation of multilevel privacy-preserving measures and highlights the implication of PUT on the former. Based on findings, the research work provides a partially hypothetical

multilevel privacy assurance graph that exhibits the privacy pathway that starts as the highest from an individual's perspective and keeps dropping in favor of inter-/intra-organizational control over PII. The privacy reaches its lowest at third-party data processing level, as shown in Figure 10. The bold font draws attention toward the contribution of this research work and four main branches of multilevel privacy assurance evaluation with recommendations for healthcare providers and respective stakeholders.

Author Contributions: Conceptualization, S.A.S., F.A.B., M.v.K.; methodology, S.A.S., F.A.B., M.v.K.; software, S.A.S.; validation, S.A.S.; formal analysis, S.A.S.; investigation, S.A.S.; resources, S.A.S., F.A.B. (arranging so that S.A.S. can do the research work); data curation, S.A.S.; writing: S.A.S., F.A.B., M.v.K.; writing—review and editing, F.A.B., M.v.K.; visualization, S.A.S.; supervision, F.A.B., M.v.K.; project administration, M.v.K.; funding acquisition, M.v.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by DMB department of EEMCS faculty of University of Twente.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All used event logs (datasets) are publicly available. See Sepsis event log [87], Hospital billing event log [88], and Urine infection event log [89].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PII	Personally Identifiable Information
PM	Process Mining
PUT	Privacy Utility Trade-off
IC	Informed Consent
HIS	Hospital Information System
EHR	Electronic Health Record
BDA	Big Data Analytics
FACT	Fair, Accurate, Confidential, Transparent
FAIR	Findable, Accessible, Interoperable, Reusable
GDP	Gross Domestic Product
DP	Differential Privacy
MPC	Materialized Privacy Claim
IM	Insurance Model
DPO	Data Protection Officer
GGD	De Gemeentelijke Gezondheidsdienst; Dutch public health group
REA	Resource Event Agent
EHBO	Eerst Hulp Bij Ongelukken (Dutch hospitals emergency service)
NZA	Nederlandse Zorgautoriteit (NZA)
IGJ	Inspectie Gezondheidszorg/Jeugd
MEB	Medicine Evaluation Board, works under Dutch Ministry of Health, Welfare and Sport (VWS)
ISO	International Organization for Standardization(ISO), Comprises 164 National standard bodies including NEN is an NGO
ISO 27001	International Standard for ISMS
ISMS	Information Security Management System: Patients' safety and patients privacy (i.e. Availability, Integrity, Confidentiality, Accountability)
NEN	Royal Netherlands Standardization Institute, Its a Non Governmental Organization (NGO)
NEN 7510 (12, 13)	Dutch Standard for ISMS
PbD	Privacy by Design
TiH	Transparency in Healthcare
TTP	Trusted Third Party

References

1. Cowie, M.R.; Blomster, J.I.; Curtis, L.H.; Duclaux, S.; Ford, I.; Fritz, F.; Goldman, S.; Janmohamed, S.; Kreuzer, J.; Leenay, M.; et al. Electronic health records to facilitate clinical research. *Clin. Res. Cardiol.* **2017**, *106*, 1–9. [CrossRef]
2. Fairfield, J.; Shtein, H. Big data, big problems: Emerging issues in the ethics of data science and journalism. *J. Mass Media Ethics* **2014**, *29*, 38–51. [CrossRef]
3. van der Aalst, W.; Bichler, M.; Heinzl, A. Responsible data science. *Bus. Inf. Syst. Eng.* **2017**, *59*, 311–313. [CrossRef]
4. ResponsibleData. Available online: <https://redasci.org/> (accessed on 27 September 2021).
5. Chassie, K. A private matter [privacy in society]. *IEEE Potentials* **2001**, *20*, 26. [CrossRef]
6. GDPR. Available online: <https://gdpr-info.eu/> (accessed on 12 March 2021).
7. DUTCH-DPA. Available online: <https://autoriteitpersoonsgegevens.nl/en/about-dutch-dpa/board-dutch-dpa> (accessed on 27 September 2021).
8. NEN(ISO). Available online: <https://www.nen.nl/en/ict/digitale-ethiek-en-veiligheid/cyber-privacy/informatiebeveiliging> (accessed on 27 September 2021).
9. NZA. Available online: <https://www.nza.nl/> (accessed on 27 September 2021).
10. Ploug, T.; Holm, S. Meta consent: A flexible and autonomous way of obtaining informed consent for secondary research. *BMJ* **2015**, *350*, h2146. [CrossRef] [PubMed]
11. Rake, E.A.; van Gelder, M.M.; Grim, D.C.; Heeren, B.; Engelen, L.J.; van de Belt, T.H. Personalized consent flow in contemporary data sharing for medical research: A viewpoint. *BioMed Res. Int.* **2017**, *2017*, 7147212. [CrossRef]
12. Critselis, E. Impact of the general data protection regulation on clinical proteomics research. *PROTEOMICS—Clin. Appl.* **2019**, *13*, 1800199. [CrossRef]
13. Garattini, C.; Raffle, J.; Aisyah, D.N.; Sartain, F.; Kozlakidis, Z. Big data analytics, infectious diseases and associated ethical impacts. *Philos. Technol.* **2019**, *32*, 69–85. [CrossRef] [PubMed]
14. Cavoukian, A.; Fisher, A.; Killen, S.; Hoffman, D.A. Remote home health care technologies: How to ensure privacy? Build it in: Privacy by design. *Identity Inf. Soc.* **2010**, *3*, 363–378. [CrossRef]
15. Kung, A.; Freytag, J.C.; Kargl, F. Privacy-by-design in its applications. In Proceedings of the 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Lucca, Italy, 20–24 June 2011; pp. 1–6.
16. Spiekermann, S.; Cranor, L.F. Engineering privacy. *IEEE Trans. Softw. Eng.* **2008**, *35*, 67–82. [CrossRef]
17. Ahlfeldt, R.M.; Huvala, I. Patient safety and patient privacy when patient reading their medical records. In *International Conference on Well-Being in the Information Society*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 230–239.
18. GDPRclause24. Available online: <https://www.privacy-regulation.eu/en/article-24-responsibility-of-the-controller-GDPR.htm#:~:text=Taking%20into%20account%20the%20nature,to%20demonstrate%20that%20processing%20is> (accessed on 27 September 2021).
19. Stahl, B.C.; Timmermans, J.; Flick, C. Ethics of Emerging Information and Communication Technologies On the implementation of responsible research and innovation. *Sci. Public Policy* **2017**, *44*, 369–381.
20. Sohail, S.A. Normative and Empirical Evaluation of Privacy Utility Trade-off in Healthcare. In Proceedings of the 33rd International Conference on Advanced Information Systems Engineering CAiSE 21, Melbourne, Australia, 28 June 2021–2 July 2021.
21. van der Aalst, W.M. Responsible data science: Using event data in a “people friendly” manner. In *International Conference on Enterprise Information Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 3–28.
22. UDHR. Available online: <https://www.un.org/en/about-us/universal-declaration-of-human-rights> (accessed on 27 September 2021).
23. EURlex. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046> (accessed on 27 September 2021).
24. GDPRprivacy. Available online: <https://gdpr.eu/data-privacy/> (accessed on 27 September 2021).
25. Eurostat. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Healthcare_expenditure_statistics (accessed on 27 September 2021).
26. Galetsi, P.; Katsaliaki, K.; Kumar, S. Values, challenges and future directions of big data analytics in healthcare: A systematic review. *Soc. Sci. Med.* **2019**, *241*, 112533. [CrossRef] [PubMed]
27. Kamal, R.N.; Lindsay, S.E.; Eppler, S.L. Patients should define value in health care: A conceptual framework. *J. Hand Surg.* **2018**, *43*, 1030–1034. [CrossRef] [PubMed]
28. da Silva Reis, J.; de Alencar Silva, P.; Bukhsh, F.A.; De Castro, A.F. Configuring Value Networks based on Subjective Business Values. In Proceedings of the 12th International Workshop on Value Modeling and Business Ontologies, VMBO 2018, Amsterdam, The Netherlands, 26–27 February 2018; pp. 158–170.
29. Emanuel, E. Shared decision making to improve care and reduce costs. *N. Engl. J. Med.* **2013**, *368*, 6–8.
30. Butzer, J.F.; Kozlowski, A.J.; Virva, R. Measuring value in postacute care. *Arch. Phys. Med. Rehabil.* **2019**, *100*, 990–994. [CrossRef]
31. Latimer, T.; Roscamp, J.; Papanikitas, A. Patient-centredness and consumerism in healthcare: An ideological mess. *J. R. Soc. Med.* **2017**, *110*, 425–427. [CrossRef]
32. Stahl, B.C.; Wright, D. Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Secur. Priv.* **2018**, *16*, 26–33. [CrossRef]

33. Pramanik, M.I.; Lau, R.Y.; Hossain, M.S.; Rahoman, M.M.; Debnath, S.K.; Rashed, M.G.; Uddin, M.Z. Privacy preserving big data analytics: A critical analysis of state-of-the-art. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, e1387. Available online: <https://wires.onlinelibrary.wiley.com/doi/epdf/10.1002/widm.1387> (accessed on 8 May 2021).
34. Sankar, L.; Rajagopalan, S.R.; Poor, H.V. Utility-privacy tradeoffs in databases: An information-theoretic approach. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 838–852. [\[CrossRef\]](#)
35. Sohail, S.A.; Krabbe, J.; de Alencar Silva, P.; Bukhsh, F.A. Privacy Value Modeling: A Gateway To Ethical Big Data Handling. In Proceedings of the 14th International Workshop on Value Modelling and Business Ontologies, VMBO 2020, Brussels, Belgium, 16–17 January 2020; pp. 5–15.
36. Erdemir, E.; Dragotti, P.L.; Gunduz, D. Privacy-Aware Time-Series Data Sharing with Deep Reinforcement Learning. *arXiv* **2020**, arXiv:2003.02685.
37. Guan, Z.; Lv, Z.; Du, X.; Wu, L.; Guizani, M. Achieving data utility-privacy tradeoff in Internet of medical things: A machine learning approach. *Future Gener. Comput. Syst.* **2019**, *98*, 60–68. [\[CrossRef\]](#)
38. Kim, K.J.; Joukov, N. *Information Science and Applications (ICISA) 2016*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 376.
39. Hlávka, J.P. Security, privacy, and information-sharing aspects of healthcare artificial intelligence. In *Artificial Intelligence in Healthcare*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 235–270.
40. McSherry, F.D. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, Providence, RI, USA, 29 June–2 July 2009; pp. 19–30.
41. Mannhardt, F.; Koschmider, A.; Baracaldo, N.; Weidlich, M.; Michael, J. Privacy-preserving process mining. *Bus. Inf. Syst. Eng.* **2019**, *61*, 595–614. [\[CrossRef\]](#)
42. Mans, R.S.; Van der Aalst, W.M.; Vanwersch, R.J. *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes*; Springer: Berlin/Heidelberg, Germany, 2015.
43. Pika, A.; Wynn, M.T.; Budiono, S.; Ter Hofstede, A.H.; van der Aalst, W.M.; Reijers, H.A. Privacy-preserving process mining in healthcare. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1612. [\[CrossRef\]](#)
44. von Voigt, S.N.; Fahrenkrog-Petersen, S.A.; Janssen, D.; Koschmider, A.; Tschorsch, F.; Mannhardt, F.; Landsiedel, O.; Weidlich, M. Quantifying the Re-identification Risk of Event Logs for Process Mining. In *International Conference on Advanced Information Systems Engineering*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 252–267.
45. Van Der Aalst, W. Process mining: Overview and opportunities. *ACM Trans. Manag. Inf. Syst. (TMIS)* **2012**, *3*, 1–17. [\[CrossRef\]](#)
46. Marazza, F.; Bukhsh, F.A.; Vijlbrief, O.; Geerdink, J.; Pathak, S.; van Keulen, M.; Seifert, C. Comparing Process Models for Patient Populations: Application in Breast Cancer Care. In *International Conference on Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 496–507.
47. Vanderfeesten, I.; Cardoso, J.; Mendling, J.; Reijers, H.A.; van der Aalst, W.M. Quality metrics for business process models. *BPM Work. Handb.* **2007**, *144*, 179–190.
48. Blum, F. Metrics in Process Discovery. Technical Report, TR/DCC. 2015; pp. 1–21. Available online: <https://www.semanticscholar.org/paper/Metrics-in-process-discovery-Blum/5542c60a14129f05baa361a5710524525cb14c17#citing-papers> (accessed on 8 May 2021).
49. Pegoraro, M.; Shankara, M.B.; van der Aalst, W.M.; Martin, L.; Marx, G. Analyzing Medical Data with Process Mining: A COVID-19 Case Study. Available online: https://www.researchgate.net/publication/351699172_Analyzing_Medical_Data_with_Process_Mining_a_COVID-19_Case_Study (accessed on 8 May 2021).
50. Dorofeev, A.; Kurganov, V.; Fillipova, N.; Pashkova, T. Ensuring the integrity of transportation and logistics during the COVID-19 pandemic. *Transp. Res. Procedia* **2020**, *50*, 96–105. [\[CrossRef\]](#)
51. Celonis. Available online: <https://www.celonis.com/blog/driving-process-inefficiencies-in-a-post-covid-economy/> (accessed on 12 October 2021).
52. Privacy-aware-pm. Available online: <https://pa.win.tue.nl/privacy-aware-process-mining/> (accessed on 12 March 2021).
53. Van Der Aalst, W. Data science in action. In *Process Mining*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 3–23.
54. Rafiei, M.; van der Aalst, W.M. Privacy-Preserving Data Publishing in Process Mining. In *International Conference on Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 122–138.
55. Dwork, C. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19.
56. Dwork, C.; Roth, A. The algorithmic foundations of differential privacy. *Found Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [\[CrossRef\]](#)
57. Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 308–318.
58. Wilson, R.J.; Zhang, C.Y.; Lam, W.; Desfontaines, D.; Simmons-Marengo, D.; Gipson, B. Differentially private sql with bounded user contribution. *Proc. Priv. Enhancing Technol.* **2020**, *2020*, 230–250. [\[CrossRef\]](#)
59. Rodríguez-García, M.; Batet, M.; Sánchez, D. A semantic framework for noise addition with nominal data. *Knowl.-Based Syst.* **2017**, *122*, 103–118. [\[CrossRef\]](#)
60. Tensorflow. Available online: <https://blog.tensorflow.org/2019/03/introducing-tensorflow-privacy-learning.html> (accessed on 12 March 2021).

61. Aldeen, Y.A.A.S.; Salleh, M.; Razzaque, M.A. A comprehensive review on privacy preserving data mining. *SpringerPlus* **2015**, *4*, 694. [CrossRef]
62. Tucker, K.; Branson, J.; Dilleen, M.; Hollis, S.; Loughlin, P.; Nixon, M.J.; Williams, Z. Protecting patient privacy when sharing patient-level data from clinical trials. *BMC Med Res. Methodol.* **2016**, *16*, 77. [CrossRef]
63. Promtool. Available online: <https://www.promtools.org/doku.php> (accessed on 12 March 2021).
64. DISCO. Available online: <https://fluxicon.com/disco/> (accessed on 27 September 2021).
65. data4TU. Available online: <https://data.4tu.nl/> (accessed on 12 March 2021).
66. Hruby, P. *Model-Driven Design Using Business Patterns*; Springer Science & Business Media: Basel, Switzerland, 2006.
67. e³ Value Modeling. Available online: <https://research.e3value.com/tools/> (accessed on 27 September 2021).
68. Avelino, J.G.; de Alencar Silva, P.; Bukhsh, F.A. Towards Green Value Network Modeling: A Case from the Agribusiness Sector in Brazil. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 458–475.
69. Simell, B.A.; Törnwall, O.M.; Hämäläinen, I.; Wichmann, H.E.; Anton, G.; Brennan, P.; Bouvard, L.; Slimani, N.; Moskal, A.; Gunter, M.; et al. Transnational access to large prospective cohorts in Europe: Current trends and unmet needs. *New Biotechnol.* **2019**, *49*, 98–103. [CrossRef] [PubMed]
70. Abedjan, Z.; Boujemaa, N.; Campbell, S.; Casla, P.; Chatterjea, S.; Consoli, S.; Costa-Soria, C.; Czech, P.; Despenic, M.; Garattini, C.; et al. Data science in healthcare: Benefits, challenges and opportunities. In *Data Science for Healthcare*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 3–38.
71. O'Connor, Y.; Rowan, W.; Lynch, L.; Heavin, C. Privacy by design: Informed consent and internet of things for smart health. *Procedia Comput. Sci.* **2017**, *113*, 653–658. [CrossRef]
72. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]
73. Christiansen, E.K.; Skjopenes, E.; Hausken, M.F.; Skeie, S.; Østbye, T.; Iversen, M.M. Shared electronic health record systems: Key legal and security challenges. *J. Diabetes Sci. Technol.* **2017**, *11*, 1234–1239. [CrossRef]
74. Mron. Available online: <https://www.mron.nl/> (accessed on 27 September 2021).
75. Medlon. Available online: <https://www.medlon.nl/> (accessed on 27 September 2021).
76. Williams, J.; Aitken, R. The service-dominant logic of marketing and marketing ethics. *J. Bus. Ethics* **2011**, *102*, 439–454. [CrossRef]
77. Eisenhardt, K.M. Agency theory: An assessment and review. *Acad. Manag. Rev.* **1989**, *14*, 57–74. [CrossRef]
78. Tihealthcare. Available online: <https://www.tihealthcare.nl/en> (accessed on 27 September 2021).
79. Recital26. Available online: <https://gdpr-info.eu/?s=recital+26> (accessed on 12 March 2021).
80. HAGA. Available online: <https://www.dutchnews.nl/news/2019/07/hospital-fined-e460000-for-privacy-breaches-after-barbie-case/> (accessed on 27 September 2021).
81. Menzis. Available online: https://gdprhub.eu/AP_-_Health_insurance_companies_Menzis_and_VGZ (accessed on 27 September 2021).
82. Mivule, K. Utilizing noise addition for data privacy, an overview. *arXiv* **2013**, arXiv:1309.3958.
83. InductiveMiner. Available online: <https://www.futurelearn.com/info/courses/process-mining/0/steps/15642> (accessed on 12 June 2021).
84. Buijs, J.C.; van Dongen, B.F.; van der Aalst, W.M. Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. *Int. J. Coop. Inf. Syst.* **2014**, *23*, 1440001. [CrossRef]
85. CoronIT. Available online: <https://nltimes.nl/2021/01/28/private-data-leak-ggd-covid-system-existed-months-report> (accessed on 12 March 2021).
86. OLGv. Available online: https://edpb.europa.eu/news/national-news/2021/dutch-dpa-fines-olvg-hospital-inadequate-protection-medical-records_en (accessed on 12 March 2021).
87. Sepsisdataset. Available online: https://data.4tu.nl/articles/dataset/Sepsis_Cases_-_Event_Log/12707639 (accessed on 12 March 2021).
88. HospitalBilling. Available online: https://data.4tu.nl/articles/dataset/Hospital_Billing_-_Event_Log/12705113 (accessed on 12 March 2021).
89. Urineinfection. Available online: https://data.4tu.nl/articles/dataset/Urineweginfectie_UWI-casus_logboek/12826727 (accessed on 12 March 2021).