# Identifying Materialized Privacy Claims of Clinical-Care Metadata Share using Process-Mining and REA ontology

Syeda Amna Sohail[1][0000−0001−8078−0411], Faiza Allah Bukhsh[1][0000−0001−5978−2754], Maurice van Keulen[1][0000−0003−2436−1372], and Johannes Gerardus Krabbe[2][0000−0003−1585−9304]

[1] University of Twente, 7522NB Enschede, The Netherlands
s.a.sohail@utwente.nl, f.a.bukhsh@utwente.nl, m.vankeulen@utwente.nl
https://www.utwente.nl/en/eemcs/dmb/
[2] Medische Spectrum Twente, Medlon BV, 7500 KA Enschede, The Netherlands
j.krabbe@mst.nl
mst.nl

**Abstract.** Metadata formation, maintenance, and interoperability are crucial for long-term, effective usage of valuable digital information across domains. Metadata interoperability especially triggers privacy concerns regarding personally identifiable information of data subjects when sensitive clinical-care metadata is shared amongst multiple caregivers. The problem intensifies when the care metadata share across caregivers is considered essentially significant for an efficient care system. Patients' un-anonymized care metadata share across caregivers is validated using a real-world Sepsis dataset with Process Mining discovery techniques. Findings are further evaluated, for both horizontally and vertically distributed caregivers, by an IT expert from a Dutch hospital. The Resource, Event, Agent (REA) ontology-based 'Insurance Model' is used to identify the underlying economic factors behind the un-anonymized patients' metadata share amongst caregivers. The model discovers the key economic agents, their prime interactions (from contract signing to the exchange of resources) for mutual economic gain/loss in the care metadata share landscape. Lastly, we explicate that the privacy concerns of patient's metadata share emerge as 'Materialized Privacy Claim'. The privacy claim only emerges if either the patient or any other potent (involved) authority finds an imbalance between the materialization and settlement of the patient's exchanged resources. The 'Materialized Privacy Claim' illustrates concretely with money claims for unlawful disclosure of a patient's personal information from caregivers and insurers.

**Keywords:** metadata share · REA ontology · Process Mining · clinical-care · privacy.

## 1 Introduction

In the contemporary digital information landscape, the formation and maintenance of metadata are vital steps to avoid information loss across domains

over time. Metadata adds context to the raw data and facilitates the extraction of knowledgeable value [3]. The metadata formation (record formulation in the repository using consistent identifiers and publication) to its maintenance (for efficient reusability) is a shared duty of the concerned authorities [23]. Additionally, metadata interoperability is the successful reuse and exchange of data sources within and across organizations Information Systems (ISs) [18]. Therefore, a collective need is felt by policymakers, regulators, and enterprises alike for metadata formation, maintenance, and interoperability within and across domains for collective cost and time-efficient growth [3, 18, 23]. Moreover, domain-specific pressure groups play a significant role in ensuring metadata share across domains [1]. Simultaneously, the local and pan-European authorities legally constrain the concerned authorities for privacy-preservation/privacy of Personally Identifiable Information (PII) of EU citizens to avoid information harm/misuse [6, 8, 10]. Privacy (literally) implies an individual's right to autonomous decision making (about what to share and with whom) and direct or indirect control over the extended personal information share [30]. With privacy-preserving measures, the legislative and regulatory authorities aim to protect the EU citizen's PII to avoid discriminatory or harmful treatment [6, 8, 10]. In this regard, special attention is given to the sensitive healthcare metadata share [10].

This research work concerns the privacy-preservation of clinical-care metadata share amongst Dutch caregivers. Here, the clinical-care is the immediate healthcare (treatment and testing) of patients [2]. (Clinical) care metadata is presumably shared un-anonymized amongst horizontally and vertically distributed caregivers in the Netherlands. Horizontally distributed caregivers are intra-organizational, internally-located caregivers (i.e. the caregivers within General Practitioners (GPs) clinic, diagnostic lab, or between various departments within a hospital) sharing care metadata. Vertically distributed caregivers are the inter-organizational, remotely located caregivers (i.e. multiple outpatient caregivers such as GPs, diagnostic labs, pharmacies, dentists, and hospitals) sharing care metadata. All these care metadata interactions involve the metadata sharing in the 'as is' condition. This implies that the data is either un-anonymized or recorded as pseudonymized data. In 'pseudonymization' the pseudo identifiers are allocated to the patients and caregivers with reversible one-way cryptography. Unlike 'anonymization', where identifiers are permanently removed for the sake of privacy [20]. In this research work, the term 'un-anonymized metadata share' is used because the patients' identities are retractable in caregivers' ISs for an efficient clinical care system [22].

The **objective and *contribution*** of this research work are:

- To validate the patients' un-anonymized care metadata share amongst Dutch caregivers using Process Mining (PM) discovery techniques on a real-world Sepsis dataset.

- To conceptualize the Dutch care metadata share landscape from a patient's perspective using REA ontology to highlight: the key economic agents, their prime interactions, and collective economic value gain or loss.

- To explicate privacy as a *'Materialized Privacy Claim (MPC)'* using REA

ontology elicited from real-world events.

To validate the patients' un-anonymized care metadata share, a real-world Sepsis dataset (extracted from a Dutch Hospital Information System (HIS) [4]) is analyzed using PM 'discovery' techniques [5,14]. PM is business process analytics on event logs (extracted directly from an organization's IS) for process discovery and compliance checking for an organization's operational improvements [31–33]. The PM results are further evaluated by an Information Technology (IT) expert working in a Dutch hospital. The Resource, Event, Agent (REA) ontology is used to locate the underlying economic factors behind patients' un-anonymized care metadata share amongst Dutch caregivers [24]. Mainly because the fundamental REA concepts are domain-independent and do not require architectural changes [24].

This paper is structured in a way that the introduction is in Section 1. Related work is given in Section 2. Section 3 contains two subsections. In subsection 3.1, the conceptual validation using a real-world Sepsis dataset with PM discovery techniques and an evaluation by an IT expert working in a Dutch hospital is given. In subsection 3.2, the REA ontology's Insurance Model (IM) is used to discover the underlying economic factors behind the current Dutch care metadata share landscape. The conclusion is in Section 4.

## 2 Related Work

In the Dutch care system, the hub and spoke model essentially facilitates care metadata formation, maintenance, and interoperability by contributing to the Electronic Health Record (EHR) from the grass-root level. Collective EHR from Dutch caregivers is publicly regulated for an uninterrupted care metadata share at the national and international level [13]. Interestingly, Dutch clinical care is appraised as the best in the EU for more than a decade till 2017 [7]. Generally, a hub, central IS, is the central point of information access for the smaller distinct reporting spokes, ISs. Hub leads the spokes to make well informed, timely, and evidence-based decisions regarding patients' treatments. The hub and spoke model ensures an efficient, patient-friendly care system with satisfied and cognitively connected caregivers [16, 25, 27]. Hub and spoke model works in both: horizontally and vertically distributed caregivers in the Netherlands for patients' primary and secondary care [16, 22, 25, 27].

Reidentification concerns of pseudonymized care metadata are already well known in academia and industry alike [15, 30]. A bigger point of concern is the lop-sided advancement of digital health technologies than their privacy-preserving measures [21]. A similar imbalance is visible in care metadata sharing efforts in comparison to their privacy-preserving measures implications [9,15,17, 26]. A leading point of concern is the open accessibility of patients' PII to numerous caregivers who are not directly involved with the patients' care [9, 17]. Such access points are principal threats to patient's sense of physical, informational, and decisional security (i.e privacy [19, 28]) [11, 12]. In addition to a patient's physical security, informational security is patients' PII's protection

from potential information harm. Decisional security is the protection of his/her autonomous decision-making regarding his/her extended PII share. Privacy-preserving measures ensure patient's informational, decisional, and physical security by the concerned authorities [19,28]. This research work addresses patients' this sense of security during the patients' clinical care.

## 3 Patients' Metadata Share and Privacy-Preservation: Conceptual Validation and REA Ontology

To validate the patients' un-anonymized metadata share amongst Dutch caregivers, the Sepsis dataset is analyzed using PM tools: Disco and ProM Lite [5,14]. Later, the PM experiments are validated by an IT expert from a Dutch hospital. Afterward, the REA's Insurance Model (IM) [24] is used to conceptualize our proposed model. The model discovered that primarily who, how, and what leads to patients' un-anonymized metadata-share amongst Dutch caregivers and in explicating privacy as a materialized claim.

### 3.1 'Sepsis' Dataset Analysis using Process Mining and Expert's Opinion

The dataset/event log comprises Sepsis patients as 'cases', treatments as 'activities' in the 'events' with initiating and ending timestamps [4]. The event log also provides the activities' link to sub-hospital organizations/departments (horizontally distributed caregivers). The **goal** of PM experiments was to look-out for evidence of patients' un-anonymized data sharing. The aim was to uncover the pairs of subsequent un-anonymized activities as indications of the lack of privacy-preserving actions.

**PM experiments findings and discussion:** The commercial tool **Disco** (fuzzy minor algorithm) discovered that there are 16 activities for 1,050 cases and 15,214 events/instances. The sub-hospital departments (with pseudo-identifiers) are shown with activities share (in percentage) in the top left box and the process model (with sharing time stamps) is given underneath see Fig. 1. The periods (with activities' median and least time stamps) of care data share amongst horizontally distributed caregivers, and activity frequency is noted. The succeeding data sharing was done either instantly or after a short period (see on arrows) from one sub-organization/department to the other. PM experiments explained that no privacy-preserving actions could practically have occurred in such brief time frames see Fig. 1.

**ProM Lite** with the social network algorithm discovers information regarding working (medical/administrative) staff in an event log. The absence of a social network suggested that either the staff disagreed to publicly share their PII or it was intentionally withheld for privacy's sake. However, the 'Dotted Chart' substantiated that the maximum succeeding activities are performed either instantly or within 2 days (48 hours) between horizontally distributed caregivers. Thus, the chart further validated our PM experiments goal see Fig. 2.
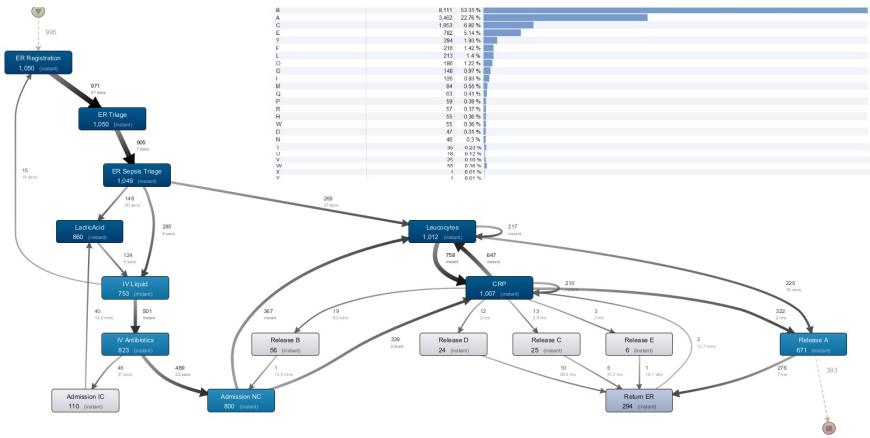
Fig. 1: Sepsis process model (Disco): patients' unanonymized care metadata share amongst horizontally distributed caregivers.
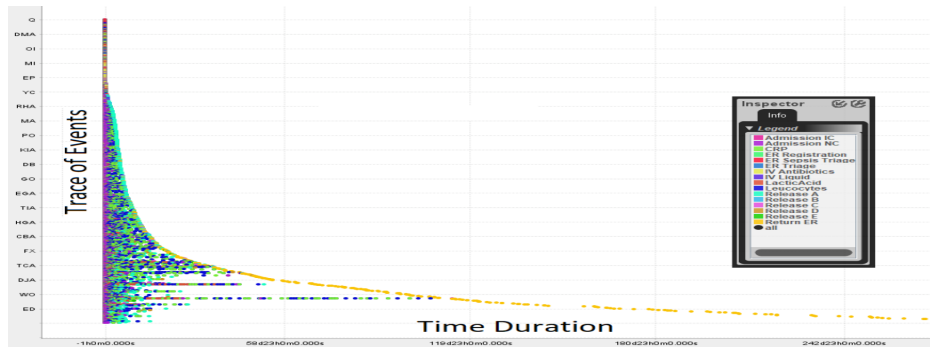


Fig. 2: Dotted Chart using ProM Lite

**Expert Opinion:** IT expert working in a Dutch local hospital was shown the PM results and was asked whether caregivers share un-anonymized (and pseudonymized) care data amongst horizontally distributed caregivers. And 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.

The REA ontology is used to conceptualize the underlying economic factors behind patients' un-anonymized metadata share amongst Dutch caregivers.

### 3.2 Underlying Conceptual Framework using REA Ontology

Dutch caregivers are privately run and partially publicly funded enterprises. Our **goal** behind using REA ontology was to discover the underlying financial prior-

ities of care enterprises for patients' un-anonymized metadata share [24]. 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 Fig. 3 conceptualization fundamentals, it is vital to emphasize that the PM experiments in Fig. 1 and Fig. 2 validated our assumptions regarding patients' un-anonymized metadata share amongst Dutch caregivers. Whereas Fig. 3 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. Fig. 3 is explained from top to bottom, describing the concepts and relations of the model. 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 like cash, metadata, and care services are exchanged as increment and/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. There can be passive 'party' agents who activate (like caregivers with respective registrations/contracts) later with further increment/decrement events. Initially, the patients and health insurers perform the increment and decrement events by exchanging cash with and for one another respectively. With an insurance contract, the PII of the insured is also stored in insurers IS. This information intake (economic resource) leaves the insured with less control over his/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 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 if there are 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 [6]) claims for the assurance of 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 Euros 460,000 for a Dutch celebrity's privacy breach [11] and charged Menzis (health insurer) Euros 50K for care data mishandling [12].

**Privacy and metadata in REA ontology:** In REA ontology the metadata formation, maintenance, and interoperability is part of an organization's 'posting and dimension aspect of financial disbursement' instead of its application
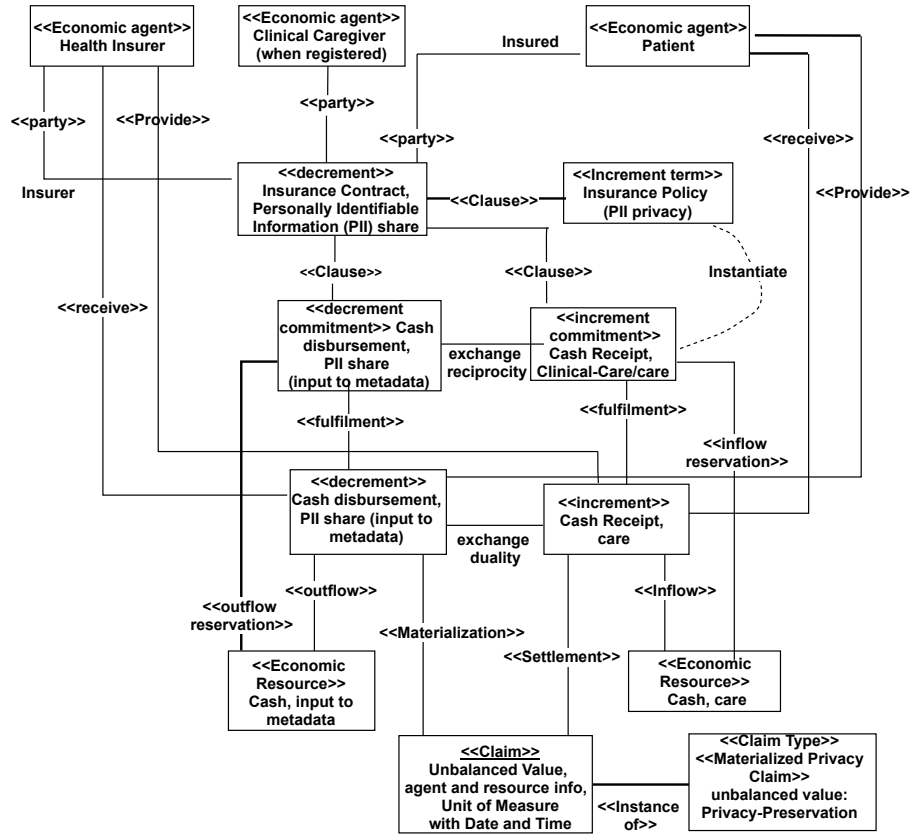
Fig. 3: REA ontology, 'Insurance Model (IM)' and 'Materialized Privacy Claim'

model [24]. Earlier (in 2006), metadata handling was considered dependent upon the behavioral pattern of the respective organization [24]. 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 [29]. Nowadays, metadata privacy is protected by the national and European regulations and involves legal commitments by the concerned authorities to avoid hefty money claims [6,8,10]. The legal and administrative requirements for privacy include privacy by policy, privacy by design, and patients' informed consent measures [6, 8, 10, 29]. 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, it is emphasized that the possibilities of these privacy claims getting compensated with hefty cash payments by the concerned authorities (caregivers/insurers) are high.

## 4    Conclusion

A real-world Sepsis dataset has been 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 (and pseudonymized/retractable) patients' metadata. An IT expert (working in a Dutch hospital) further evaluated the PM results and their generalizability for 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. In this regard, the REA model helped us to discover: who, how, and what leads to an un-anonymized metadata-share amongst vertically and horizontally distributed Dutch caregivers.

REA's Insurence 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). As a decrement, the patient pays monthly payments to the insurer and *his/her PII* as an input to the metadata of the insurer and caregivers. The decremented resources become valuable to the receiver and let him/her evaluate the imbalance (if there is any) between the materialization of economic resources (provided care for the physical, informational, and decisional security) and their prior settlements (cash payments, contract clauses ensuring privacy). An important result of our conceptualization is that the privacy concerns take the form of a Materialized Privacy Claim (MPC) 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.

The future work includes the breaking down of the proposed REA model into two sub-models with further specifications and detailed descriptions of mutual transactions between patients, health insurers, and caregivers in binary format.

Process Mining, REA ontology and clinical-care 'Materialized Privacy Claim'

# References

1. aalep, `http://www.aalep.eu/lobbying-landscape-netherlands`, 7 Jan, 2021
2. Collinsdictionary, `https://www.collinsdictionary.com/dictionary/english/clinical-care`, 11 Nov, 2020
3. Ctis, `https://www.ema.europa.eu/en/documents/newsletter/clinical-trials-information-system-ctis-highlights-june-2020_.pdf`, 20 Dec, 2020
4. data.4tu, `https://data.4tu.nl/`, 11 Nov, 2020
5. Disco, `https://fluxicon.com/disco`, 11 Nov, 2020
6. Dpo, `https://www.itgovernance.eu/nl-nl/data-protection-officer-dpo-under-the-gdpr-nl`, 11 Nov, 2020
7. Echi-2017, `https://healthpowerhouse.com/media/EHCI-2016/EHCI-2016-launch-presentation.pdf`, 11 Nov, 2020
8. Edpb, `https://edpb.europa.eu/`, 11 Nov, 2020
9. Euctregister, `https://www.clinicaltrialsregister.eu/ctr-search/trial/2017-002976-24/NL`, 11 Nov, 2020
10. Gdpr, `https://gdpr-info.eu/`, 11 Nov, 2020
11. Hagahospital, `https://www.cordemeyerslager.nl/fine-of-e-460-000-imposed-on-dutch-haga-hospital-by-dutch-data-protection-officer-the-first-dutch-fine-under-gdpr-july-19-2019/`, 30 Dec, 2020
12. insurerfine, `https://www.iapp.org/news/a/dutch-dpa-hits-medical-insurer-with-50k-euro-gdpr-fine/#:~:text=The%20Dutch%20data%20protection%20authority,its%20processing%20of%20personal%20data.`, 30 Dec, 2020
13. Nictiz, `https://www.nictiz.nl/english/exchange-of-electronic-patient-data-in-the-netherlands/the-infrastructure-for-central-exchange/#section1`, 30 Dec, 2020
14. promtool, `https://www.promtools.org/doku.php`, 11 Nov, 2020
15. Wiley, `https://www.wiley.com/en-us/Medical+Information+Systems+Ethics-p-9781848218598`, 11 Nov, 2020
16. Albarello, F., Prati, F., Sangiorgi, L., Tremosini, M., Menegatti, M., Depolo, M., Rubini, M.: Does hub-and-spoke organization of healthcare system promote workers' satisfaction? Journal of Applied Social Psychology **49**(10), 634–646 (2019)
17. Badawy, R., Hameed, F., Bataille, L., Little, M.A., Claes, K., Saria, S., Cedarbaum, J.M., Stephenson, D., Neville, J., Maetzler, W., et al.: Metadata concepts for advancing the use of digital health technologies in clinical research. Digital biomarkers **3**(3), 116–132 (2019)
18. Bountouri, L., Papatheodorou, C., Soulikias, V., Stratis, M.: Metadata interoperability in public sector information. Journal of Information Science **35**(2), 204–231 (2009)
19. Chassie, K.: A private matter [privacy in society]. IEEE Potentials **20**(4), 26 (2001)
20. Critselis, E.: Impact of the general data protection regulation on clinical proteomics research. PROTEOMICS–Clinical Applications **13**(2), 1800199 (2019)
21. Davis, J.S., Osoba, O.: Improving privacy preservation policy in the modern information age. Health and Technology **9**(1), 65–75 (2019)
22. Garattini, C., Raffle, J., Aisyah, D.N., Sartain, F., Kozlakidis, Z.: Big data analytics, infectious diseases and associated ethical impacts. Philosophy & technology **32**(1), 69–85 (2019)
23. Habermann, T.: Metadata and reuse: Antidotes to information entropy. Patterns **1**(1), 100004 (2020)

24. Hruby, P.: Model-driven design using business patterns. Springer Science & Business Media (2006)
25. Joseph, P.: Eliminating disparities and implicit bias in health care delivery by utilizing a hub-and-spoke model. Research Ideas and Outcomes **4**, e26370 (2018)
26. Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M., Michael, J.: Privacy-preserving process mining. Business & Information Systems Engineering **61**(5), 595–614 (2019)
27. Nocera, N.: Hubs, spokes and trauma nurse coordinators: New south wales' model of optimal trauma care—part i. Australian Emergency Nursing Journal **6**(1), 5–9 (2003)
28. Serour, G.: Confidentiality, privacy and security of patients' health care information. International Journal of Gynecology and Obstetrics **2**(93), 184–186 (2006)
29. Sohail, S.A., Krabbe, J., de Alencar Silva, P., Bukhsh, F.A.: Privacy value modeling: A gateway to ethical big data handling. In: 14th International Workshop on Value Modelling and Business Ontologies, VMBO 2020. pp. 5–15. CEUR (2020)
30. Van Den Hoven, J.: Information technology, privacy, and the protection of personal data. Information technology and moral philosophy pp. 301–322 (2008)
31. Van Der Aalst, W.: Process mining: Overview and opportunities. ACM Transactions on Management Information Systems (TMIS) **3**(2), 1–17 (2012)
32. Van Der Aalst, W.: Data science in action. In: Process mining, pp. 3–23. Springer (2016)
33. Vanderfeesten, I., Cardoso, J., Mendling, J., Reijers, H.A., van der Aalst, W.M.: Quality metrics for business process models. BPM and Workflow handbook **144**, 179–190 (2007)