Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 29, S. 144-149, Aralık 2021

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 29, pp. 144-149, December 2021

Research Article

Enriching the Open Provenance Model for a Privacy-Aware Provenance Management

Dilek Yılmazer Demirel[1*], Özgü Can[2]

[1*] Istanbul Technical University, Faculty of Computer and Informatics Engineering, Department of Computer Engineering, İstanbul, Turkey, (ORCID: 0000-0002-4008-4478), demirel18@itu.edu.tr

[2] Ege University, Faculty of Engineering, Department of Computer Engineering, İzmir, Turkey, (ORCID: 0000-0002-8064-2905), ozgu.can@ege.edu.tr

(International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2021 – 21-23 October 2021)

(**DOI:** 10.31590/ejosat.1023420)

ATIF/REFERENCE: Yılmazer Demirel, D., & Can, O. (2021). Enriching the Open Provenance Model for a Privacy-Aware Provenance Management. *European Journal of Science and Technology*, (29), 144-149.

Abstract

Today, the total amount of data that is generated, copied, and stored are increasing rapidly. Thereupon, the trustworthiness of the data source and the quality of data have significant importance for an effective data analysis. Therefore, it is critical to improve accountability for the quality of data. For this purpose, provenance information is used to provide the quality of data. Provenance information ensures the reliability and quality of data. Data provenance is a form of metadata to describe the life cycle of a data. Therefore, provenance information maintains the history of the data by describing how data are derived. The Open Provenance Model (OPM) aims to meet the requirements of a provenance model. For this purpose, OPM defines a core set of rules. Thus, OPM provides provenance interoperability. In this study, OPM is enhanced to provide a Privacy-Aware Provenance Management (PAPM) model. The goal of the PAPM model is to use provenance information in order to protect data from unwanted access and detect security violations. Therefore, PAPM uses provenance information to protect data privacy. Since the proposed PAPM model is domain-independent, it can be integrated into any interested domain to preserve privacy and ensure data security.

Keywords: Provenance, Open Provenance Model, Privacy, Data Security, Knowledge Engineering, Semantic Web.

Mahremiyet-Farkında Bir Köken Yönetimi için Açık Köken Modelinin Zenginleştirilmesi

Öz

Günümüzde üretilen, kopyalanan ve depolanan toplam veri miktarı hızla artmaktadır. Bunun sonucu olarak, etkin bir veri analizi için veri kaynağının güvenilirliği ve verinin kalitesi büyük önem taşımaktadır. Bu nedenle, veri kalitesi için izlenebilirliği arttırmak çok önemlidir. Bu amaçla, veri kalitesini sağlamak için köken bilgisi kullanılmaktadır. Köken bilgisi, verilerin güvenilirliğini ve kalitesini sağlamaktadır. Veri kökeni, verinin yaşam döngüsünü tanımlayan bir meta veri biçimidir. Bu nedenle, köken bilgisi, verilerin nasıl türetildiğini açıklayarak verilerin geçmişini korumaktadır. Açık Köken Modeli (OPM), bir köken modelinin gereksinimlerini karşılamayı hedefklemektedir. Bu amaçla, OPM temel bir kurallar kümesi tanımlamaktadır. Böylelikle, OPM köken birlikte çalışabilirliğini sağlamaktadır. Bu çalışmada, Gizlilik-Farkında bir Köken Yönetimi (PAPM) modeli sağlamak için OPM genişletilmiştir. PAPM modelinin amacı, verileri istenmeyen erişimlerden korumak ve güvenlik ihlallerini tespit etmek için köken bilgisini kullanmaktır. Bu nedenle PAPM, veri mahremiyetini korumak için köken bilgisini kullanmaktadır. Önerilen PAPM modeli etki alanından bağımsız olduğundan, mahremiyeti korumak ve veri güvenliğini sağlamak için herhangi bir etki alanına entegre edilebilecektir.

Anahtar Kelimeler: Köken, Açık Köken Modeli, Mahremiyet, Veri Güvenliği, Bilgi Mühendisliği, Anlamsal Web.

[*] Corresponding Author: demirel18@itu.edu.tr

1. Introduction

Today's information technologies produce a massive volume of data and need to keep track of the origin of data and the metadata to support data accountability and to improve data quality. Provenance is the detailed information about the origin of data and the history of operations made on data. In (Omitola et al., 2010), the provenance term is described as "*Provenance, also known as lineage, describes how an object came to be in its present state, and thus, it describes the evolution of the object over time*". Provenance does not directly enforce the information security requirements, but it provides the evidence to support the data security (Phua et al., 2018). For this purpose, provenance keeps track of how and where the data was generated, steps that were performed on the processing of data, and by whom these operations were performed. Thus, provenance addresses the data accountability issue and helps to find out whether a policy violation or a data breach has taken place (Tan et al., 2015).

The Open Provenance Model (OPM) is a community data model for provenance that facilitates the meaningful interchange of provenance information between systems (Kwasnikowska et al., 2015). The OPM allows to characterize how things are dependent on others and resulted in specific states and expresses these dependencies with a directed graph (Moreau et al., 2011). Provenance information has no common representation and infrastructure. Therefore, the comprehensibility of provenance information and data exchange becomes very difficult. For this purpose, OPM provides an interoperability layer to express provenance information from different systems in a digitally represented form and allows provenance information to be exchanged between these systems. Consequently, provenance information in different systems is represented in a more meaningful way.

The Semantic Web which is described as an extension of the current web uses ontologies to provide a shared and common understanding of a specific domain. Semantic Web represents information in a machine-understandable and machine-processable format. Therefore, interoperability between systems is supported. As stated in (Golbeck & Hendler, 2008), Semantic Web is a natural fit to represent the provenance information.

In this study, a Semantic Web based Privacy-Aware Provenance Management (PAPM) model is proposed. The presented model aims to track all changes that are made on data since its first creation to its current state. Also, the model allows to access data according to the defined access permissions and restrictions, and to detect security violations by tracking the provenance information. Thus, the main focus of the proposed model is ensuring data privacy by using provenance information. For this purpose, the PAPM model is based on the OPM and generated by utilizing the OPM Profile for Dublin Core (DC Profile). The Dublin Core (DC) terms (Dublin Core Terms, 2021) are metadata about resources and OPM Profile for Dublin Core maps provenance related Dublin Core metadata terms to the OPM and allows existing Dublin Core provenance to be re-expressed in OPM (Miles et al., 2009). The subject of this study is enhancing the OPM model by adding new concepts within the scope of the proposed PAPM model. In this study, the enhanced OPM model and the conceptual view of the PAPM model are presented.

The structure of the paper is organized as follows. The literature review is presented in Section 2. Section 3 clarifies the PAPM model with the extensions on the OPM model. Finally, Section 4 concludes and describes the future work.

2. Related Work

Data provenance has been studied with respect to various fields, such as database systems, digital libraries, art, archaeology, and workflow management systems. The research challenges and the application of provenance in the metadata of digital libraries are discussed in (Burgess, 2016). In (Davidson & Freire, 2008), opportunities and challenges in scientific workflows and provenance are discussed. A formal provenance model to specify control-flow driven scientific workflows is proposed in (Butt & Fitch, 2021). Also, a provenance model named ProvOne+ for scientific workflows is presented and validated for the agricultural domain in (Butt & Fitch, 2021). Moreover, provenance is studied for IoT environments. The requirements, challenges, and applications of data provenance in the IoT are explored in (Butt & Fitch, 2020).

Provenance also supports data accountability, data integration, and data quality. In (Tan et al., 2013), a survey on security and data accountability in distributed systems is presented. A provenance based solution for the cloud data accountability is discussed in (Tan et al., 2015). An analysis framework is presented in (Cheahi & Plale, 2013) to detect conflicts and ambiguities in provenance traces by identifying errors that occur in the provenance processing. A data provenance model to support instance level data integration processes is proposed in (Tomazela et al., 2013).

In recent years, provenance studies are focused on blockchain based researches. In (Suhail et al., 2020), a blockchain based provenance framework for product traceability is proposed. The use of blockchain to record supply chain provenances in a trustworthy manner is studied in (Garrard et al., 2020). A blockchain-based trusted cloud data provenance architecture is proposed in (Liang et al., 2017).

In this study, different from the existing studies, provenance information is used to track access violations. Therefore, the OPM model is enhanced for the Privacy-Aware Provenance Management model. The goal of the PAPM model is to detect privacy threats based on provenance information. This study is based on the provenance model presented in (Can & Yilmazer, 2014; Can & Yilmazer, 2020). The formal representation of the model is given (Can & Yilmazer, 2014). In (Can & Yilmazer, 2020), a provenance model to integrate the provenance and security concepts in order to detect privacy violations is presented. Also, in (Can & Yilmazer, 2020), the related model is demonstrated for the health care domain to preserve patients' privacy. On the other hand, distinct from these studies, this study focuses on the extensions that are performed on the OPM model.

3. Material and Method

3.1. Open Provenance Model (OPM)

Provenance is the documented history of an object. The dictionary definition of provenance is *the place of origin or earliest known history of something* (Provenance, 2021). It is also defined as the documentation of processes in a digital object's life cycle and accepted as a crucial component of workflow systems (Moreau et al.,). The Open Provenance Model (OPM) is a model of provenance that defines provenance in a precise manner and

supports a digital representation of provenance. The OPM uses Provenance Interoperability Layer to allow provenance information to be exchanged between different systems. The OPM assumes that the provenance of objects (whether digital or not) is represented by a directed acyclic graph which is enriched with annotations that capture further information pertaining to execution (Moreau et al., 2011). Therefore, the OPM provides a common representation for provenance information. Figure 1 and Figure 2 show a system before the OPM and after the OPM (Moreau et al., 2008), respectively. Before the OPM-based systems, the provenance data is stored distribute in different systems.

The OPM is based on `Artifact`, `Process`, and `Agent` nodes. The `Process` is an action that takes an `Artifact` object as an input and creates a new `Artifact` object as output. The `Agent` represents the subject that performs the action which is the `Process`. The `Artifact` represents a physical object or a digital representation of an object in any of the states that change during the runtime. Therefore, a notation is needed to indicate these representations of the same resource in different states. Thus, one `Artifact` is a version of another `Artifact`. Different profile definitions have emerged in order to customize the OPM for any domain. The OPM profile aims to use Dublin Core (DC) terms on concepts of OPM. DC terms are metadata about resources. In the OPM, resources do not communicate directly. Hence, relationships between resources should be defined and these relations should be associated. Therefore, DC terms are used. The provenance-related DC metadata terms map to OPM graphs to allow existing DC provenance to be re-expressed in OPM (Dublin Core Terms, 2021). The aim of this mapping is to be connected to wider provenance information available in OPM data.

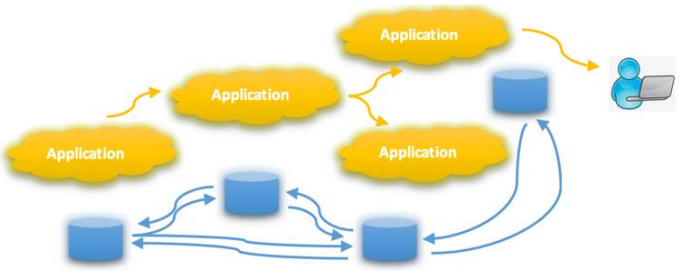

Figure 1. Before OPM

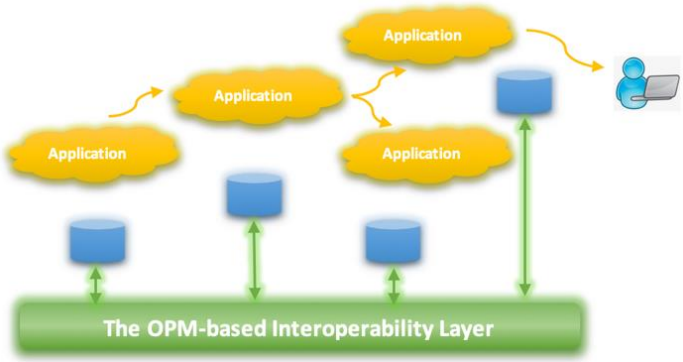

Figure 2.After OPM

Table 1. The DC terms and their functionality on OPM concepts.

Dublin Core/OPM Terms	Functionality on OPM Concepts
`dc:isVersionOf`	Indicates the version relationship between resources.
`dc:hasVersion`	Indicates the version of a resource.
`dc:creator`	Represents the person, organization, or system that is the creator of a resource.
`dc:date`	Indicates a date or time period associated with a process in a resource's lifecycle.
`dc:contributor`	Represents the person, organization, or system that contributed to the resource in order to create the next version of the resource.
`dc:accrualMethod`	Indicates the method that is used to add items to a resource collection.
`dc:available`	Indicates the date range that the resource became or will become available.
`wasDerivedFrom`	Indicates resources that are derived from each other.
`wasGeneratedBy`	Specifies the relationship between a resource and a process.
`wasControlledBy`	Specifies the relationship between a process and a system or a person.

4. Results and Discussion

4.1. Enrichment of the OPM for the PAPM Model

The proposed PAPM model aims to support security and privacy by controlling access to personal information and preventing the unwanted disclosure of personal information. The provenance information determines the origin of the data and allows to track operations that are performed on the data. Therefore, the provenance concept is used to perform access control and to trace authorizations. For this purpose, the PAPM model is developed based on the OPM. While developing the PAPM model, some terms of the OPM are used directly and some terms are used indirectly. Classes that are defined in the PAPM model are as follows: `Subject`, `Person`, `Object`, `Artifact`, `Organization`, `Operation`, `Service`, `Permission`, `Prohibition`, `Authorization`, `Right`, `AccessHistory`, `Role`, `Purpose`, `Gender`, `MaritalStatus`. The relationship between the concepts of OPM and PAPM is shown in Table 2.

Table 2. The relationship between the concepts of OPM and PAPM..

Class Name	OPM	PAPM
Subject	dct:Agent dct:creator dct:contributor dct:Publisher	A person, an organization, or a service
Person	dct:Agent dct:creator dct:contributor dct:publisher	The subclass of Subject
Organization	dct:Agent dct:creator dct:contributor dct:publisher	The subclass of Subject
Service	dct:Agent dct:creator dct:contributor dct:publisher	The subclass of Subject
Artifact	dct:Collection dct:BibliographicResource	Personal information that could identify an individual
Operation	dct:instructionalMethod	Actions performed in the interested domain
Authorization	No term	Species authorizations for whether or not Artifact objects can be accessed
Prohibition	No term	The subclass of Authorization that defines access prohibitions
Permission	No term	The subclass of Authorization that defines access permissions
Right	dct:rights	The set of rights
Purpose	No term	The access purpose condition that is used by determining authorizations
Role	No term	The role of each user in the interested domain (such as doctor, patient, etc.)
AccessHistory	No term	The detailed access information that is performed on the Artifact object (by whom, when and for what purpose the access is performed)
Object	No term	The general set for objects that could be accessed in the interested domain
Gender	No term	The gender information for the Person
MaritalStatus	No term	The marital status information for the Person

The Subject class refers to the actor who performs operations such as creating a new data, adding or modifying an existing record. The object properties that are defined for the Subject class are as follows: hasRole, isOwnerOf, requests and performs. These properties do not exist in the OPM. Thus, they are specifically defined in the PAPM model. hasRole defines the role of the Subject, isOwnerOf is the owner of the personal information, requests indicates the operation that the Subject wants to perform and performs states the operation that the Subject is performed.

The Person class is the subclass of the Subject class and refers to a real person. The Person class has object and data properties to define demographic data and relations between persons. These properties are also specified in the PAPM model and do not exist in the OPM. personMaritalStatus, personGender, hasFather, hasMother and hasChild are the object properties defined for the Person class; the data properties are personFirstName, personSecondName, personSurname, personFullName, personAge, personBirthdate, personIdentityNumber, personPhoneNumber, personEmail and personAddress.

The Organization class refers to the organization that performs operations on an object. organizationName and organizationDescription are data properties defined for the Organization class. While organizationName is stated in the OPM with the dc:title term, organizationDescription is defined in the PAPM and states the description information for the Organization.

The Service class indicates a service that performs operations on an object. The data properties of the Service class are serviceName and serviceDescription. Similar to the Organization class, serviceName is specified in the OPM with the term dc:title and serviceDescription is defined in the PAPM and states the description information for the Service.

The Artifact class refers to data that is given as input to an operation or received as an output after the operation. artifactName, artifactValue, hasFormat, isVersionOf, hasVersion, priorVersion, Format, versionInfo, Creator, Created, Available, dateCopyrighted, hasOwner, hasAccessHistory, hasPermission, hasProhibition, isInputOf and isOutputOf are the properties defined for the Artifact class. Within the context of OPM, properties and their equivalent terms in OPM are: artifactName (dct:title), hasFormat (dct:hasFormat), isVersionOf (dct:isVersionOf), hasVersion (dct:hasVersion), priorVersion (dct:priorVersion), Format (dct:format), versionInfo (dct:versionInfo), Creator (dct:creator), Created (dct:created), Available (dct:available), dateCopyrighted (dct:dateCopyrighted). artifactValue (the value of the stored Artifact data), hasAccessHistory (links the AccessHistory to an Artifact when the Subject accesses to the Artifact), hasPermission (links an Artifact to a Permission to define access permissions), hasProhibition (links an Artifact to a Prohibition to define access prohibitions), hasOwner (links an Artifact to the owner of the personel information), isInputOf (links an Artifact as an input for the Operation), and isOutputOf (links an Artifact as a result of an Operation) are concepts that are added for the PAPM model.

The Operation class states an operation such as adding, updating, or deleting that is performed by an actor. The Operation class has the following object and data type properties: operationName, date, creator, created, operationPerformedDate, hasOutput, operationMinApplicableAge, isPerformedBy, operationMaxApplicableAge, isRequestedBy, operationPeriodForLowRisk, hasInput, operationPeriodForHighRisk, isPerformedFor. Properties and their equivalent terms in OPM are as follows: operationName, (dct:title), date (dct:date), creator (dct:creator), created (dct:created) and type (dct:type). Properties that are added in the scope of this study are as follows: operationPerformedDate (the date that the operation is performed), operationMinApplicableAge (indicates the minimum age if an age limit is required for the operation to be performed), operationMaxApplicableAge (indicates the maximum age if an age limit is required for the operation to be performed), operationPeriodForLowRisk (states the time interval of the operation that will be applied for low risk groups), operationPeriodForHighRisk (states the time interval of the operation that will be applied for high risk groups), hasInput (states the data that the operation receives as input), hasOutput (states the data that the operation produces as output), isPerformedBy (indicates the person who performed the operation), isPerformedFor (indicates for what the operation is performed) and isRequestedBy (states the person who requested the operation).

The Permission class defines the permit accessibility of an Artifact. On the contrary, the Prohibition class defines the access restriction of an Artifact. Both classes has the relevant properties: expiredDate, creator, created, permissionName/prohibitionName, hasPurpose, hasRelatedSubject, hasRight, startedDate, modified, contributor. permissionName/prohibitionName (dct:title), creator (dct:creator), created (dct:created), modified (dct:modified) and contributor (dct:contributor) are the properties that have equivalent terms in OPM. The new properties added for the Permission and the Prohibition are hasPurpose (defines the purpose condition that the permission/prohibition will be valid), hasRelatedSubject (the person or role that the permission/prohibition is associated with), hasRight (indicates the right that the data can be accessed/cannot be accessed, such as read, write, etc.), startedDate (indicates the date that the validity of the permission/prohibition starts) and expiredDate (indicates the date that the validity of the permission/prohibition ends).

The Purpose class expresses the purpose for accessing the data. This class defines the purpose condition while defining permissions and prohibitions. The purposeName, purposeDescription, creator, created, modified, and contributor are properties that are defined for the Purpose class. In these properties, purposeDescription (specifies the description for the purpose) is added for the PAPM model. The rest of the properties have equivalent terms in the OPM. The Role class states the role of a Subject in the interested domain. The properties of the class are roleName, roleDescription, creator, created, modified, and contributor. roleDescription that specifies the description for the role is added to the PAPM model, while other properties have equivalent terms in the OPM. The Right class refers to the rights that are used for the data access, such as read, write, and etc. The rightName, rightDescription, creator, created, modified, and contributor are properties of the Right class. The OperationType class indicates the type of operations that are defined. The properties of this class are creator, created, operationTypeName, operationTypeDescription, modified, and contributor. For both of these classes, operationTypeDescription and rightDescription are new properties that specify the description for the access right and operation type, respectively. Finally, the Gender and MaritalStatus classes and their properties are related with the Person class. The genderName and maritalStatus are defined for the related classes and the equivalent term for these properties in the OPM is dct:title.

Consequently, the access history for data can be fetched by enriching the OPM model to preserve privacy. Also, versions of data can be tracked to monitor access violations. Therefore, the

extended model provides to trace the data from the moment it is first created and to store details of all access information related to data. As the final extended model is domain-independent, it could be used in any domain to prevent the unauthorised accesses and to provide the protection of sensitive data.

4. Conclusions and Recommendations

Provenance specifies the origin of data and provides information about the evolution of data. Therefore, provenance information improves the data quality and trustworthiness of data by providing information about data from its first creation. In this study, provenance is studied in the scope of information security and data privacy. Thus, a Privacy-Aware Provenance Management model is presented to protect data from unwanted access and detect security violations. The presented model is based on the OPM model. In the scope of this study, the OPM model is enhanced for the PAPM model and details of the related extensions are presented. As future work, a blockchain-based approach will be integrated into the PAPM model to provide a tamper-proof information and a generic PAPM framework will be implemented. Also, this generic blockchain-based PAPM framework will be evaluated to support security and privacy. For this purpose, a use case study for preserving organizational privacy will be demonstrated and the blockchain-based PAPM framework will be validated.

References

Burgess, L.C. (2016). Provenance in Digital Libraries: Source, Context, Value and Trust. In: Lemieux V. (eds) Building Trust inInformation. Springer Proceedings in Business and Economics, pp. 81-91. Springer, Cham.

Butt, A.S, & Fitch, P. (2021). A provenance model for control-flow driven scientific workflows. Data & Knowledge Engineering, 131-132, 101877.

Butt A.S., & Fitch P. (2020). ProvONE+: A Provenance Model for Scientific Workflows. In: Huang Z., Beek W., Wang H., Zhou R., Zhang Y. (eds) Web Information Systems Engineering – WISE 2020. Lecture Notes in Computer Science, Vol 12343, pp. 431-444. Springer, Cham.

Can, O., & Yilmazer, D. (2014). A Privacy-Aware Semantic Model for Provenance Management. In: Closs S., Studer R., Garoufallou E., Sicilia MA. (eds) Metadata and Semantics Research (MTSR 2014). CCIS Vol 478, pp. 162-169. Springer, Cham.

Can, O., & Yilmazer, D. (2020). A novel approach to provenance management for privacy preservation. Journal of Information Science, 46(2):147-160.

Can, O., & Yilmazer, D. (2020). Improving privacy in health care with an ontology-based provenance management system. Expert Systems, 37(1), 12427.

Cheahi Y.W., & Plale, B. (2014). Provenance quality assessment methodology and framework. ACM Journal of Data and Information Quality, 5(3), Article 9.

Davidson, S.B., & Freire, J. (2008). Provenance and scientific workflows: challenges and opportunities. In: Proceedings of the 2008 ACM SIGMOD international conference on management of data (SIGMOD'08), pp. 1345-1350.

Dublin Core Terms, https://www.dublincore.org/specifications/dublin-core. Last accessed 14 Nov 2021.

Elkhodr, M., & Mufti, Z.B. (2019). On the challenges of data provenance in the Internet of Things. International Journal of Wireless & Mobile Networks (IJWMN), 11(3):43-52.

Garrard, R., & Fielke, S. (2020). Blockchain for trustworthy provenances: A case study in the Australian aquaculture industry. Technology in Society, 62, 101298.

Golbeck, J., & Hendler, J. (2008). A semantic web approach to the provenance challenge. Concurrency and Computation: Practice and Experience, 20(5): 431-439.

Kwasnikowska, N., Moreau, L., & Van Den Bussche, J. (2015). A Formal Account of the Open Provenance Model. ACM Transactions on the Web, 9:2, Article 10.

Liang, X., et al. (2017). ProvChain: A Blockchain-based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability. In: 17th IEEE/ACM Int. Symposium on Cluster, Cloud and Grid Computing (CCGRID), pp. 468-477.

Miles, S., Moreau, L., & Futrelle, J. (2009). OPM Profile for Dublin Core Terms (Draft). https://nms.kcl.ac.uk/luc.moreau/papers/dc-opm09.pdf. Last accessed 14 Nov 2021.

Moreau, L., Clifford, B., Freire, J., et al. (2011). The Open Provenance Model core specification (v1.1). Future Generation Computer Systems, 27(6): 743-756.

Moreau L., Freire J., Futrelle J., McGrath R.E., Myers J., & Paulson P. (2008). The Open Provenance Model: An Overview. In: Freire J., et al. (eds) Provenance and Annotation of Data and Processes. IPAW 2008. LNCS, Vol 5272. Springer, Berlin, Heidelberg.

Omitola, T., Gibbins, N., & Shadbolt, N. (2010). Provenance in Linked Data Integration. Future Internet Assembly.

OPM Tutorial. Interoperability. https://openprovenance.org/opm/tutorial/slides/6-interoperability.pptx. Last accessed 14 Nov 2021.

Phua, T.W., & Ko, R.K.L. (2018). Data Provenance for Big Data Security and Accountability. In: Sakr S., Zomaya A. (eds) Encyclopedia of Big Data Technologies. Springer, Cham.

Provenance. https://www.lexico.com/definition/provenance. Last accessed 14 Nov 2021.

Suhail, S., Hussain, R., Khan, A., & Seon Hong, C. (2020). Orchestrating product provenance story: When IOTA eco system meets electronics supply chain space. Computers in Industry, 123, 103334.

Tan, A.Y.S, et al. (2015). Provenance for cloud data accountability. The Cloud Security Ecosystem Technical, Legal, Business and Management Issues. 1st Edn. Chapter 8, pp. 171--185. Syngress, MA, USA.

Tan, Y.S., Ko, R.K.L., & Holmes, G. (2013). Security and Data Accountability in Distributed Systems: A Provenance Survey. In: IEEE 10th International Conference on High Performance Computing and Communications \& 2013 IEEE International Conference on Embedded and Ubiquitous Computing, pp. 1571-1578.

Tomazela, B., Hara, C.S., Ciferri, R.R., & de Aguiar Ciferri, C.D. (2013). Empowering integration processes with data provenance. Data & Knowledge Engineering, 86:102-123.