sensors

MDPI

Systematic Review

Data Provenance in Healthcare: Approaches, Challenges, and Future Directions

Mansoor Ahmed [1,2,*], Amil Rohani Dar [2,3], Markus Helfert [1], Abid Khan [4] and Jungsuk Kim [5,6,*]

1 ADAPT Centre, Innovation Value Institute, Maynooth University, W23 F2H6 Maynooth, Ireland; markus.helfert@mu.ie
2 Department of Computer Science, COMSATS University, Federal Capital, Islamabad 44000, Pakistan; amil.rohani@uokajk.edu.pk
3 Department of Computer Science & Information Technology, Faculty of Computing & Engineering, University of Kotli, Azad Jammu and Kashmir, Kotli 11100, Pakistan
4 College of Science and Engineering, University of Derby, Derby DE22 1GB, UK; a.khan3@derby.ac.uk
5 Department of Biomedical Engineering, Gachon University, Seongnam-si 13120, Republic of Korea
6 Research Institute, Cellico Company, Seongnam-si 13449, Republic of Korea
* Correspondence: mansoor.ahmed@mu.ie (M.A.); jungsuk@bme.gachon.ac.kr (J.K.)

Abstract: Data provenance means recording data origins and the history of data generation and processing. In healthcare, data provenance is one of the essential processes that make it possible to track the sources and reasons behind any problem with a user's data. With the emergence of the General Data Protection Regulation (GDPR), data provenance in healthcare systems should be implemented to give users more control over data. This SLR studies the impacts of data provenance in healthcare and GDPR-compliance-based data provenance through a systematic review of peer-reviewed articles. The SLR discusses the technologies used to achieve data provenance and various methodologies to achieve data provenance. We then explore different technologies that are applied in the healthcare domain and how they achieve data provenance. In the end, we have identified key research gaps followed by future research directions.

Keywords: data provenance; healthcare; provenance technologies; cryptography; ontologies; blockchain

Citation: Ahmed, M.; Dar, A.R.; Helfert, M.; Khan, A.; Kim, J. Data Provenance in Healthcare: Approaches, Challenges, and Future Directions. *Sensors* **2023**, *23*, 6495. https://doi.org/10.3390/s23146495

Academic Editor: Isabel De la Torre Díez

Received: 26 April 2023
Revised: 3 July 2023
Accepted: 12 July 2023
Published: 18 July 2023

1. Introduction

The definition of data provenance was introduced by authors [1] many years ago. Various domains have been used for data provenance such as E-services, artificial intelligence, and healthcare. Data provenance is suitable for tracing data history. In healthcare services, data provenance is highly important [2]. We studied the different usage of data provenance by authors in various domains such as [3–13]. In the past, the management and recording of provenance information used to be a manual process. However, due to the sheer volume of provenance information, manual handling of these tasks has become challenging. Scientists and engineers now face significant challenges in both data management and recording provenance information, even for basic inquiries.

In existing research, data provenance is achieved with the help of different technologies. These are log-based [14,15], cryptography-based [16–18], blockchain-based [19–21], and ontology-based [22,23]. Some of the existing works are using these technologies to achieve and improve the data provenance as different research has been carried out in this direction [24–27]. We considered different healthcare applications of data provenance proposed by authors such as electronic health record sharing, personal health data, electronic mobile health applications, COVID vaccination data, and electronic patient data.

After 2018, GDPR compliance is an important aspect. There is a need to explore GDPR compliance in the existing research for data provenance in the healthcare sector. Because data provenance consists of sensitive information, it belongs to patients in the healthcare system. For EU citizens, there was a data protection act in 1995 which was replaced by

the General Data Protection Regulation (GDPR) in 2018 [28]. The purpose was to protect EU citizens' rights and to make companies comply with regulations according to GDPR rules [29]. This opens a new direction for researchers, and very little effort is made in this respect.

The purpose of the SLR is to focus on data provenance in healthcare data and perform an extensive literature review on existing techniques for achieving provenance and existing related work in the healthcare domain to find open research issues for future research directions. The aim and objective of this SLR is:

- To study the state-of-the-art technologies used for data provenance in healthcare.
- To provide insight to readers about important attributes of data provenance for healthcare applications.
- To identify the latest trend to achieve data provenance.
- To identify research gaps and provide future research directions.

We used the methodology set by Kitchenham and Charters [30] to conduct this SLR. After careful analysis of research articles, inclusion and exclusion criteria were applied. We considered research articles according to the research questions. We answer the research questions based on an in-depth analysis of these research articles. This helped us to identify potential research gaps, and researchers can provide solutions in their future research.

The rest of the paper is organized as follows: Section 2.1 provides the data provenance overview, Section 2.2 provides the healthcare overview, Section 2.3 describes healthcare and data provenance to understand the requirements of provenance in the healthcare field, and Sections 2.4 and 3 show the proposed methodology for conducting an SLR related to data provenance in healthcare. The results and discussion are provided in Sections 4 and 5. The conclusion is provided in Section 7.

2. Background

This section introduces data provenance and healthcare data. Then, the importance of data provenance in healthcare is discussed.

2.1. Data Provenance Overview

In this section, we introduce the basic concepts of data provenance, its main usages, implementation technologies, and the basic architecture of data provenance systems.

2.1.1. Data Provenance

Different authors have defined data provenance in different ways [31]. According to some researchers, it is defined as the origin or source of data [1], while others define [32] data provenance as tracing the source of data and how and from where data came. According to authors in [33–35], data provenance is defined as collecting the whole process of information from the generation and evolution of data over time. The authors in [36] defined data provenance in terms of healthcare data as follows: "Provenance is defined as attributes about the origin of health information at the time it is first created and tracks the uses and permutations of the health information over its lifecycle".

2.1.2. Usage of Data Provenance

With the help of provenance information, we can find the origin of data and the operations performed on data at different stages and times. Data provenance has been applied in different domains [34]. There are different usages of data provenance in different domains. A list of data provenance usage is provided in Table 1.

Table 1. Usages of data provenance.

Sr #	Usage	Ref.
1	Ensure data trustworthiness and data quality.	[3]
2	Detect errors in data generation and processing.	[4–6]
3	Help in data recovery and trust management between sensors.	[7]
4	Improve data readability and describe data citation.	[8]
5	Useful in forensic investigation in the field of IoT healthcare.	[9]
6	In healthcare, the complete information of healthcare creation, access, and transfer.	[10]
7	Achieving trust using data provenance for cloud provider.	[11]
8	In home care, why and how procedures applied in treatment can be improved.	[12]
9	Trusted document history of artwork.	[13]

2.1.3. Short Description of Technologies Used for Data Provenance

Currently, there are four types of technologies [34,37,38] available for data provenance. The names for these technologies and short descriptions are provided in Figure 1 and in the sub-section. These technologies can achieve data provenance with the help of each other. The authors in [39] achieved provenance in the supply chain using ontology-based blockchain systems. More details are available in the upcoming section.

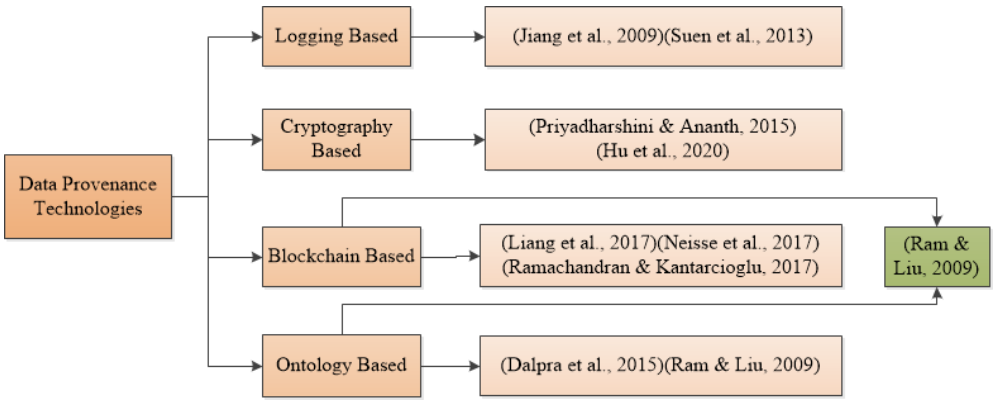

Figure 1. Data provenance technologies [14–16,20,21,23,31,34,40].

(a) Logging-based technology

Logs have pivotal significance in data provenance as they furnish an intricate account of events and activities transpiring within a system or application. Logs play a crucial role in capturing the history and lineage of data within a system or application. They provide a detailed record of events, actions, and interactions related to the data, including their creation, modification, movement, and access. The problems associated with available logs can be resolved much faster compared to those lacking attached logs [14]. Traditional log management systems store the logs in just one node. This limitation can be overcome using an end-to-end provenance mechanism as proposed by [15]. Logs contribute to data provenance through providing a comprehensive record of data-related events, enabling traceability, accountability, auditing, incident investigation, problem diagnosis, and performance optimization.

(b) Cryptography-based technology

Digital signature and Message authentication code are the cryptographic mechanisms [16]. With the help of these techniques, one can find the origin of data. The problem with cryptographic technology is that it did not provide data processing history [34]. It is important

to note that while cryptographic techniques contribute to data provenance through addressing security and integrity aspects, achieving comprehensive data provenance may require additional measures such as metadata capture, audit trails, and logging mechanisms.

(c) Blockchain-based technology

Data provenance in the blockchain can be achieved through blockchain transactions that record data operations [19]. Blockchain-based provenance examples include provchain [20]. Some authors achieved provenance using smart contracts [21,40]. These approaches help in achieving data provenance using blockchain.

(d) Ontology-based technology

In the computer science community, ontology-based research has achieved tremendous attention in the field of databases, artificial intelligence, and computational linguistics [22]. For ontology-based provenance, there exist some ontologies [23,31].

2.1.4. General Architecture for Data Provenance

The general architecture in Figure 2 explains that users (data subjects) can perform actions on data stored on the server. Whenever the user takes an action on the data, their corresponding provenance data will be stored in the provenance store. In the same fashion, whenever the data controller takes any action on the data, their corresponding provenance data will also be stored in the provenance store. Similarly, if the data controller wants to share the data with any third party, we can store the provenance data for user trust and satisfaction. The data controller can query the provenance store and may obtain the corresponding results.

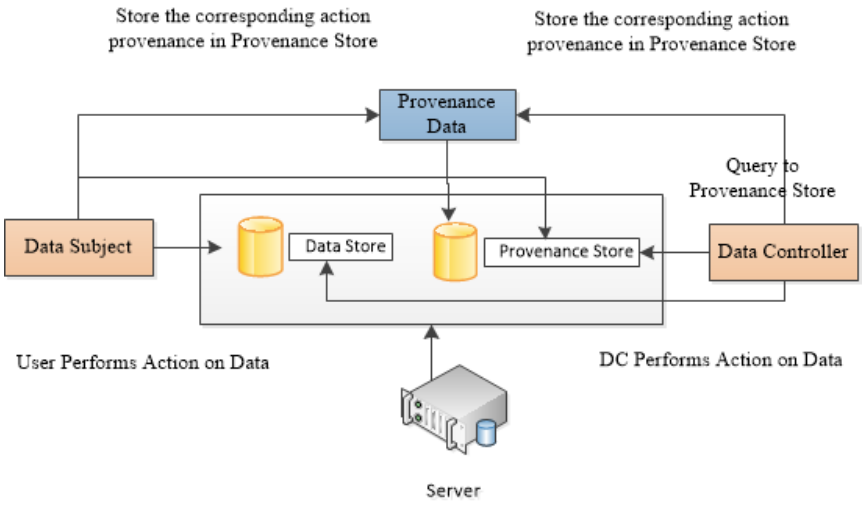

Figure 2. General data provenance architecture.

2.2. Healthcare Overview

In this section, we have provided an overview of healthcare data, the role of healthcare systems, healthcare in terms of the World Health Organization (WHO), and the typical architecture of the healthcare system.

2.2.1. Healthcare Data

Electronic health systems or digital healthcare systems consist of healthcare data and are implemented all over the world. According to the European Health Digital Service Infrastructure, patients are treated positively using healthcare data [36]. Patients have electronic health records [41] in developed and developing countries. Researchers of healthcare big data in the world are struggling to deal with the multidimensional nature of healthcare data. Also, the sharing of healthcare-sensitive data is a challenge, and data providers hesitate while doing this [42]. Researchers, analysts, patients, and doctors want

to access healthcare data instantly and data provenance in this situation becomes more important [1].

2.2.2. Healthcare Stakeholders

Currently, there are four stakeholders in healthcare data. These are patients, providers, payers, and providers [42]. Communication and collaboration among these stakeholders are of much importance to making a detailed analysis of patient medical records. These entities are linked with security and privacy threats. Patients produce data and these data are from clinical records or sensors as wearable devices [43]. Payers pay the healthcare cost, i.e., private companies or insurance companies. Medical records are collected and stored by providers. An analysis is performed by researchers and analysts on the data provided by providers to improve the performance of the healthcare industry. Figure 3 shows the relationship between the stakeholders of healthcare data.

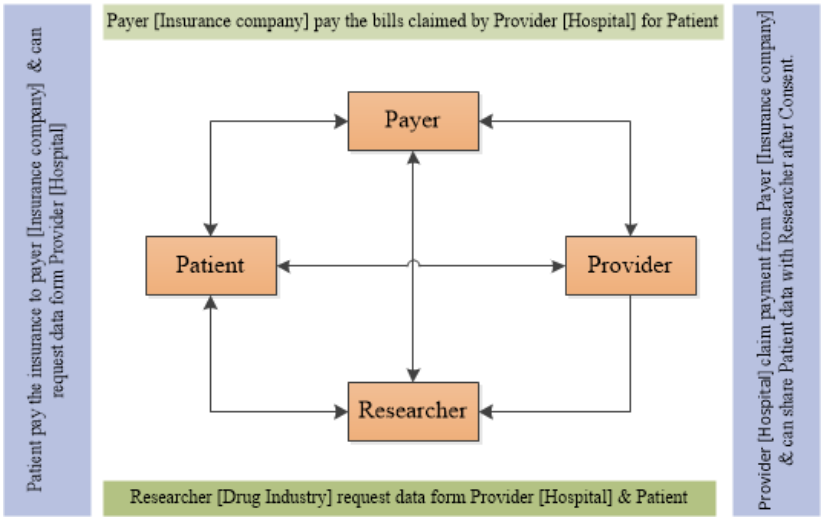

Figure 3. Healthcare data stakeholders.

There is a need for strong collaboration and communication between the stakeholders of healthcare records. Privacy and security issues are interlinked between these entities because patients are a source of data in healthcare systems. The data are produced using patient medical records and from wearable devices [43]. Payers such as insurance companies, bank lenders, etc. paid the healthcare cost directly or indirectly. Providers collect and store the patient's health records. A researcher is an entity that collects the record of a patient from a provider with patient consent or directly from the provider.

2.2.3. The WHO and Healthcare

To strengthen the healthcare system, the WHO has devised a health model in terms of six building blocks [44]. The six building blocks are service delivery, health workforce, information, medical products, vaccines and technologies, financing, and leadership/governance [45]. The six building block goals or outcomes are improved health, responsiveness, financial risk protection, and improved efficiency. The WHO's health system framework with these six building blocks helps in identifying the strengths and weaknesses of any health system in the world [44].

2.3. Healthcare and Data Provenance

In this section, we have explained the importance of data provenance in the healthcare system. Though healthcare systems can work alone with data provenance services, they can provide better health services to patients [46]. Also, an example structure of data provenance is provided at the end of this section.

Healthcare data can be used by doctors, researchers, and analysts to achieve their goals. For research purposes, electronic health data offer huge potential for researchers, and computational power is growing day by day with methodological developments that can deal with big data easily [47,48]. The data belong to the patient, and they become risky when shared with doctors, researchers, or analysts. In healthcare, adversaries are always ready to attack patient data. It is always important to record who is accessing the patient's health data, when, and why. So, provenance becomes critical for patient data safety [36]. Although electronic health record systems define how they work and how they share/exchange patient data [46], they do not discuss, capture, or store the process of information sharing. Here, the provenance system helps in storing or recording the way the data were created or shared with other entities [49]. Data provenance helps in providing the answers to questions about who collected the data, why they collected the data, when they collected the data, and what type or part of the data was collected [37].

Table 2 represents an example of a data provenance record consisting of several fields which show how data provenance may look [50]. Each action of the data subject on the data, as depicted in Figure 2, will be stored in the data provenance store with the corresponding action.

Table 2. Structure of data provenance entry.

#	ID	Date	Time	Action			Revision	Hash
				Name	Reason	Location		
1	11	25 May 2022	12:40:22	Create	Blood Test	Florida	0	X
2	12	26 May 2022	14:23:01	Append	Medical Report	Texas	1	Y
3	-	-	-	-	-	-	-	-

2.4. GDPR, Healthcare, and Data Provenance

For EU citizens, there was a data protection act in 1995 which was replaced by the General Data Protection Regulation (GDPR) in 2018 [28]. The purpose was to protect EU citizens' rights and to make companies comply with regulations according to GDPR rules [29]. Citizens are known as data subjects in GDPR terms. The GDPR guides companies on how to use and process data subject data and guides the data subject about how to use and control the data. Because any sort of data like phone number, email address, location, etc. [51] can reveal the identity of a data subject, citizens are very worried about privacy issues due to rapid changes in technology and digitalization in all sectors like healthcare. Healthcare data consist of the medical history of a patient [52]. According to GDPR Article 9, health data are a special category of data.

There is a need to develop GDPR-compliant healthcare systems with data provenance. There exists prior literature about the GDPR articles. The authors in [53,54] used a logical deletion method to solve the issue of Article 17 of the GDPR. Article 7 of the GDPR emphasizes consent management. Data subject data cannot be stored without prior consent [55] and the authors in [56] proposed a solution according to GDPR compliance. Finding healthcare systems that are complying with GDPR in terms of data provenance is also the focus of this SLR.

3. Research Methodology

First, in a systematic literature review (SLR), we have defined specific research questions. It uses a well-defined methodology to answer those questions through collecting, classifying, and extracting all existing research [57–59]. The basic steps in conducting any SLR are shown in Figure 4.

Different authors have proposed guidelines for writing a systematic literature review. However, in this SLR, we followed the steps recommended by Barbara Kitchenham [57–59].

Many high-impact-factor peer-reviewed journals [60,61] followed this methodology. This process is very famous and specifically designed for conducting systematic reviews in the field of computing research.

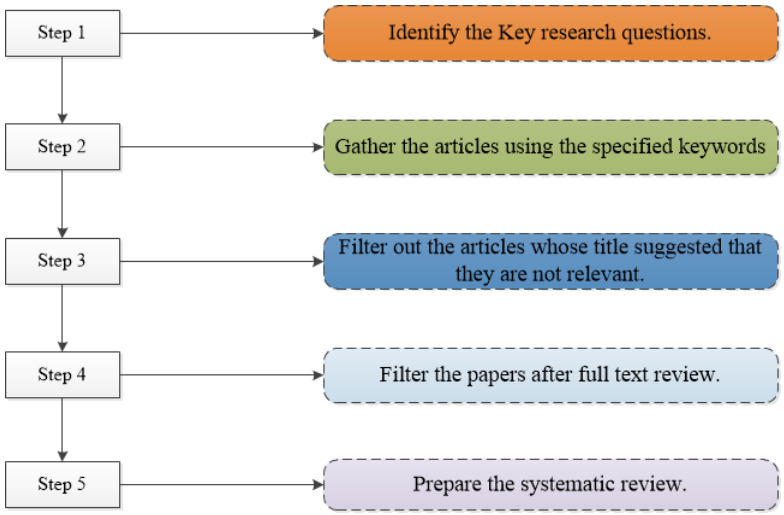

Figure 4. Basic steps in conducting the review.

3.1. Need of Conducting SLR

According to our knowledge, there is no survey on data provenance in the healthcare domain that describes how existing technologies are used to achieve data provenance. In Section 4, we briefly described the need of conducting SLR because there is always a need to store the actions of the data subject or data controller, as depicted in Figure 2, and provenance can help in various ways, as described in Table 1. Therefore, the applications of healthcare with data provenance features need to be discussed in detail.

3.2. Research Question (RQ) and Motivation

RQ-1 Which technologies were used for achieving data provenance?

RQ-2 Which combinations of technologies were used to achieve data provenance?

RQ-3 Which application in healthcare achieves privacy, security, integrity, traceability, unforgeability, and compliance using the technologies discussed in RQ-1 and RQ-2?

RQ-4 What are the major challenges and issues for future research?

Motivation for each research question is mentioned in Table 3.

Table 3. Research questions and motivation.

RQ	Motivation
1	The objective of this research question is to highlight the technologies which achieve provenance.
2	The objective of this research question is to find the technologies which achieve data provenance in combination with each other.
3	The objective of this research question is to find healthcare applications that achieve data provenance while maintaining privacy, security, integrity, traceability, unforgeability, and compliance.
4	The objective of this question is to find out those data provenance challenges and issues which need to be addressed in the future.

3.3. Search Strategy

Pursuing research questions, the following search queries were executed to collect maximum literature for review: "Data Provenance and Healthcare data", "Data Provenance and Electronic Health Records", "Blockchain-based Data Provenance", and "Data

Provenance Technologies". After careful analysis of databases (Google Scholar, ACM, and IEEE Explore), we collected 785 studies.

3.4. Inclusion and Exclusion Criteria

In this phase, the collected studies were analyzed within the research scope, i.e., data provenance and healthcare. Some studies were found exactly aligned with the research area and some were found partially or completely out of the research scope. The articles were selected based on an outline set by [57]. The following exclusion and inclusion criteria were applied to extract the final list of articles. Figure 5 shows the steps that how inclusion and exclusion criteria are applied in the form of a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) chart. We followed the steps for article selection according to the methodology set by Kitchenham in [57] and used the PRISMA chart for only this particular step of inclusion and exclusion.

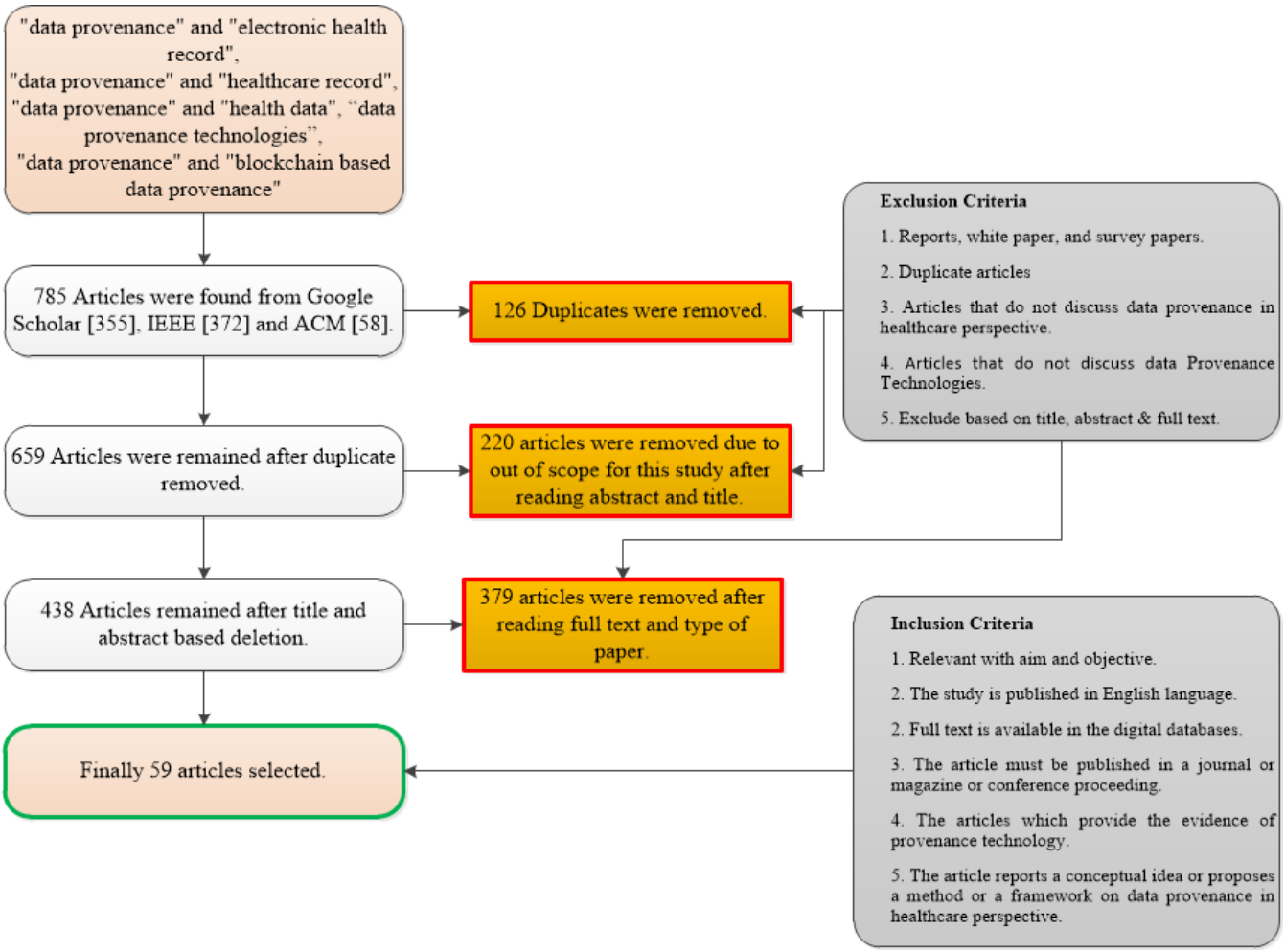

Figure 5. PRISMA chart for SLR.

Exclusion criteria:

(a) Reports, white papers, and survey papers;
(b) Duplicate articles;
(c) Articles that do not discuss data provenance from a healthcare perspective;
(d) Exclude based on title, abstract, and full text.

Inclusion criteria:

(a) Relevant with aim and objective;
(b) The study is published in the English language;

(c) Full text is available in a digital database;
(d) The article must be published in a journal or magazine or conference proceeding;
(e) The article reports a conceptual idea or proposes a method or a framework for data provenance from a healthcare perspective.

3.5. Classification Criteria

The research questions in this SLR consist of two parts to analyze the existing research for data provenance in the healthcare domain. We classified the shortlisted studies according to the research questions.

3.6. Data Extraction

In the next step, we extracted and analyzed the shortlisted studies to find out the information according to the research questions. For each included study, the data extraction tables were set and filled. Table 4 compares papers for further analysis based on the specific characteristics which are crucial for data provenance. These are discussed below.

Table 4. Articles published by various publishers.

Article Type	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Conference	2	2	3	2	5			6	2	7	3	2	
Journal			1		1	1	1	3	3	1	6	6	1
Chapter	1												
Thesis								1					

3.6.1. Technologies

For data provenance, we found four types of technologies from [34,37,38]. These can also work in combination to achieve data provenance, like in [36,39,62]. We tried to find out the trend in using these technologies for data provenance in the healthcare domain in recent times.

3.6.2. Security

A data provenance system needs to satisfy fundamental security requirements such as confidentiality, integrity [34,50], and availability [34,50,63]. It is crucial to consider the privacy implications as data provenance can potentially expose sensitive information. Therefore, encrypting both the data provenance and the source data becomes essential in ensuring the protection and privacy of the information. Data inside the provenance store must be immutable. There is a need to secure data provenance, which helps in maintaining the confidentiality and integrity of source data, and the security of the provenance is also important. The provenance of data must be available for users at any time from anywhere. In the case of patient data in an emergency case, it is subject to high availability.

3.6.3. Storage

According to the authors in [64], provenance data can be stored both ways, e.g., in centralized and distributed locations. Using a centralized approach, the maintenance is very difficult to manage, whereas in a distributed approach, the cost matters. We have explained this in the Discussion section.

3.6.4. Metadata

Provenance data are a type of metadata that belongs to an entity and records the creation and usage of data sources, also called lineage or pedigree [33,34]. Provenance metadata are of utmost importance, just like the data itself, because an adversary may use

these metadata to perform a privacy attack. Metadata, which describe the data source, demand privacy protection, especially in smart health applications. For example, a patient's disease history consists of highly sensitive information [65]. There is a need to focus on this aspect of data provenance.

3.6.5. Overhead

In any application, minimizing overhead is crucial for achieving high performance [66]. It is imperative to ensure that the impact of provenance collection remains minimal [67]. The authors in [66] considered less overhead as a design goal to increase the performance of the proposed system. Through reducing unnecessary overhead, such as computational or resource-intensive processes, the system can operate more efficiently and effectively. There must be low overhead for accessing and storing provenance data. This directly affects the scalability issue because data provenance transactions put an extra burden on the application with source data transactions. There is a need to control this cost.

3.6.6. Unforgeability

It pertains to the act of falsifying data provenance through manipulating both the source data and introducing counterfeit data. Provenance must be tightly coupled with its source data and the provenance system must detect if an adversary tries to forge any fake data. Someone can forge a medical report to avoid an investigation of misdiagnosis [50]. There is a need for such a system which has the feature of unforgeability.

3.6.7. Compliance

Various types of data, including phone numbers, email addresses, and location information, have the potential to expose the identity of individuals [51]. With rapid advancements in technology and widespread digitization across sectors like healthcare, citizens have become increasingly concerned about privacy issues. The growing integration of technology and digitalization has heightened these worries, particularly regarding the safeguarding of personal information. Healthcare data consist of the medical history of a patient [52]. There is a need to develop GDPR-compliant healthcare systems with data provenance. In this SLR, we tried to find healthcare systems that are complying with the GDPR in terms of data provenance. Article 30 of the GDPR refers to recording processing activities and Article 32 refers to the storage of processing.

4. Results

4.1. Research Trend

The findings of the systematic review are summarized in this section. Table 4 shows the number of publications per year from 2010 to 2022 considered for this study. From the table, we observed that there is 1 thesis, 1 book chapter, 24 journals, and 33 conference papers.

Figure 6 shows the distribution of papers by author country-wise. Researchers from different countries are working on the problem for which this SLR was conducted. From this figure, we can figure out that most of the affiliations of authors belong to the USA and China.

For observing research trends and impact, the best way is to find the number of publications in a year. There are many research articles published in a year by various publishers. We have identified relevant articles from these publishers according to the research aims and objectives. In Figure 7, we found the publication trend over the years using Table 4. From 2017 to 2021, the research in this direction makes progress. We have shown the importance of the topic in a timeline diagram in Section 5.

The number of articles published by various publishers and publication types are shown in Figures 8 and 9, respectively. We explored three digital libraries (Google Scholar, IEE, and ACM). From the results (as shown in Figure 9), we found that most articles are published in conferences. Most of the papers we selected for this SLR are from Google Scholar. The Google Scholar database shows the research articles published by various

publishers. But we also searched two other reliable repositories (ACM and IEEE Explore) for this SLR.

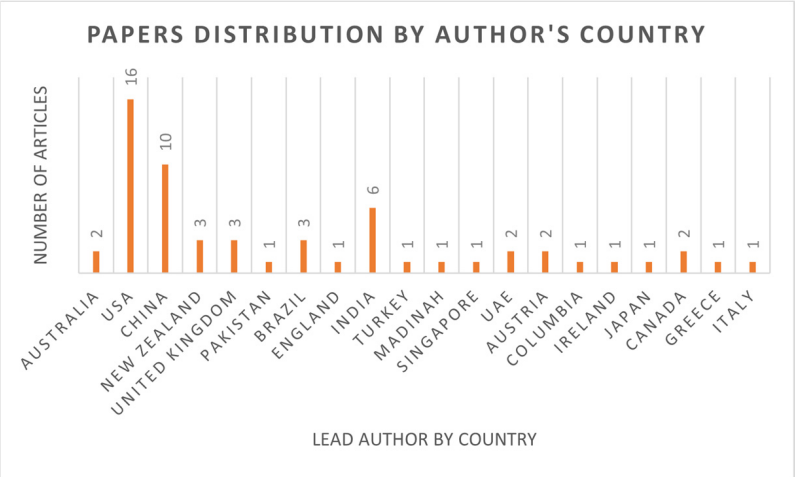

Figure 6. Authors by country.

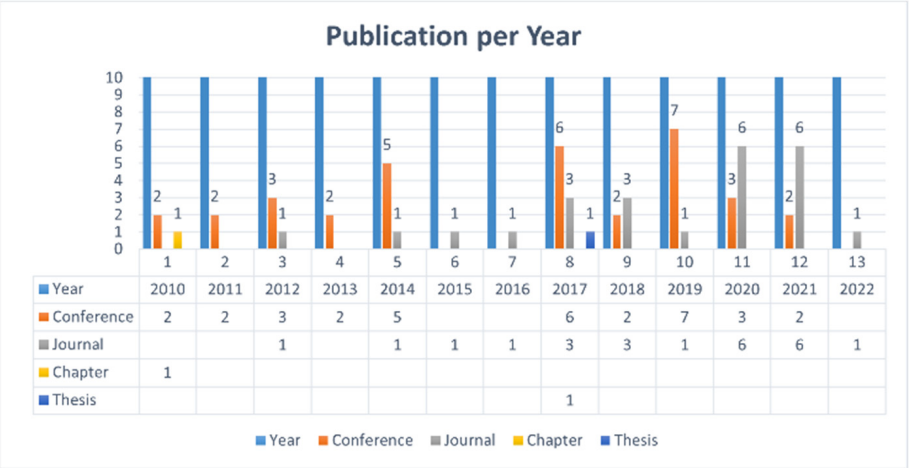

Figure 7. Publication trend over the years.

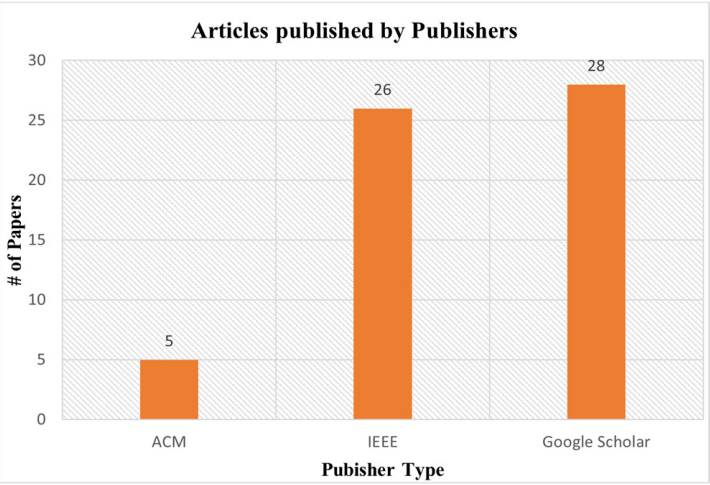

Figure 8. Articles published by publisher type.

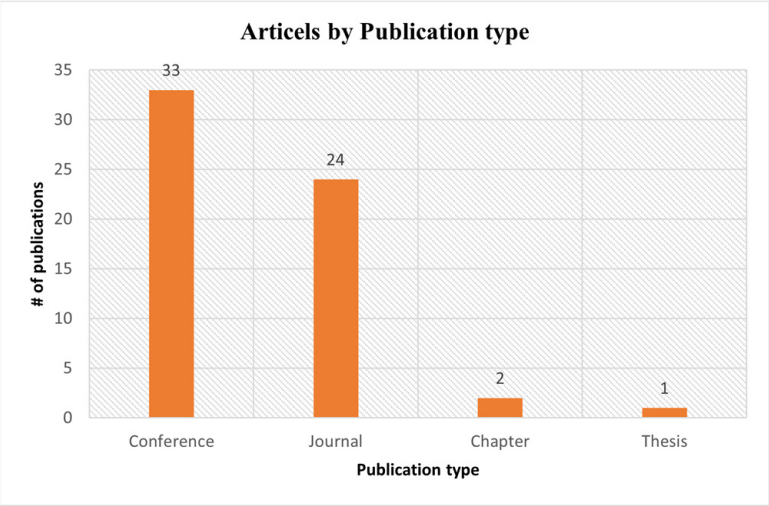

Figure 9. Articles published by various publication types.

4.2. Research Questions

In this section, we utilize our findings to cross-examine the research questions we have identified in Section 4. We provide a discussion of our analysis and address the limitations of the review.

4.2.1. Addressing RQ-1

The first research question considered for this study was "Which technologies were used for achieving data provenance?" In Section 2, Figure 1, the names of the technologies are provided with a brief description. Now, we briefly present these technologies and their purpose in Table 5. That is how different research works used these technologies.

Table 5. Data provenance techniques and their purpose in different research works.

Ref.	Provenance Technology	Purpose
[10]	Blockchain-based	The authors proposed a generic framework based on blockchain which is simplified in adoption and provides data provenance.
[15]	Log-based	The authors proposed a data event logging system in the cloud which captures, analyzes, and visualizes data events. It deals with cloud security problems related to critical data.
[16]	Cryptography-based	The authors in this work propose an approach called cryptographic provenance verification consisting of two modules: sign and verify. The purpose was to protect data from fake keystroke injection.
[17]	Cryptography-based	The author in this technique proposed a watermarking-based method for provenance embedding to avoid insider and outsider attacks.
[20]	Blockchain-based	The purpose of the authors was to achieve data provenance through detecting operations on cloud files as objects with the help of blockchain technology. They store the provenance in blockchain transactions.
[23]	Ontology-based	The authors in this work presented a proposal using the PROV process based on PROV-O ontology to capture the process data provenance in software processes to improve product quality.
[31]	Ontology-based	The authors proposed a w7 model which captures the provenance semantics for data in any domain. The purpose was to capture domain-specific provenance.

Table 5. *Cont.*

Ref.	Provenance Technology	Purpose
[68]	Log-based	Using logs, the authors proposed a model for identifying and collecting provenance data from log files and showed that capturing data provenance through logs is very helpful in some circumstances.
[69]	Log-based	According to the authors, this research deals with data security and data activity audit. It was applied to a cloud system.
[70]	Log-based	In this work, the author extracts data provenance through reconstructing log files. He models the information from log files into provenance relations.
[71]	Log-based	In this work, the authors stored provenance data as a separate file and used this provenance to detect changes in the dataset.
[72]	Log-based	The authors proposed a Prov-Trust system that stores the provenance using log files which capture the events of smart contracts or via blockchain transactions, depending on the provenance change event.
[73]	Log-based	The authors proposed a lightweight provenance tracing system based on system event logging and unit-level taint propagation.
[74]	Cryptography-based	Using advanced cryptography algorithms, the authors proposed a data provenance system for secure hosts.
[75]	Cryptography-based	In this paper, the authors proposed a lightweight novel scheme to securely transmit sensor data provenance. This scheme avoids the tampering of data from adversary attacks.
[76]	Cryptography-based	The authors in this research work provided efficient techniques for provenance encoding. This technique is based on a dynamic Bayesian network.
[77]	Cryptography-based	According to the authors, the proposed scheme reduced the storage requirement and computational time for the Internet of Things.
[78]	Cryptography-based	In this paper, the authors proposed a scheme to trace the origin and transformation history of multimedia data sharing and dissemination.
[79]	Blockchain-based	The authors in this research achieve data provenance and data integrity using blockchain with smart contracts.
[80]	Blockchain-based	The authors identified the functional and non-functional requirements for a secure data provenance framework in IoT. They also proposed a solution for scalability and privacy issues in the blockchain.
[81]	Blockchain-based	The author proposed cloud-based data provenance using blockchain. This work traces operations on data and generates provenance. Due to the global blockchain network, tampering with provenance is challenging.
[82]	Blockchain-based	The authors proposed a blockchain-based approach with smart contracts for drug traceability in healthcare.
[83]	Blockchain-based	The author proposed a blockchain-based model for data provenance. Also, their purpose was the security and traceability of personal health data with data provenance.
[84]	Ontology-based	The proposed system detects privacy violation to reduce privacy risks in the healthcare domain. It was implemented using semantic technologies.
[85]	Ontology-based	The purpose was to provide a security model for protecting data provenance using semantic web technologies.
[86]	Ontology-based	This research work deals with the source of data published on the web using provenance vocabulary.
[87]	Ontology-based	In this work, the authors proposed the Provenir ontology. It was designed for the origin of genetic data and has similar features to provenance vocabulary [86].
[88,89]	Blockchain-based	According to the authors, the origins of assets are traceable (like patient records) using blockchain technology and records can be confirmed.

4.2.2. Addressing RQ-2

The second research question considered for this study concerned "the combination of technologies that were used for achieving data provenance". In Section 2, Figure 1, an example is given in the figure which shows the combination of technologies, e.g., blockchain and ontology-based combination, for achieving the data provenance. Here, we have presented a combination of technologies and their purpose in Table 6 that explains how different research works use these technologies and achieve their goal. The purpose of this table is to bring the attention of researchers toward achieving data provenance using different combinations of technologies.

Table 6. Combinations of technologies to achieve data provenance.

Ref.	Combination of Technologies	Purpose
[24]	Blockchain- and cryptography-based	The purpose of the author in this research is to store the provenance data efficiently using blockchain and IPFS [25] technology with the secure hash function SHA-256.
[26]	Blockchain- and ontology-based	The author used blockchain and PROV ontology specifications for software provenance.
[27]	Blockchain- and ontology-based	The author of this paper used the ontology and blockchain combination to control the flow of personal data. The use of ontology helps in identifying the entities involved in personal data processing.
[36]	Blockchain and ontology-based	In this research work, the authors solved the problem of data provenance using blockchain and PROV ontology.
[39]	Blockchain- and ontology-based	In this paper, to determine the provenance, the author used the traceability ontology and enforced traceability constraints on platforms based on the Ethereum blockchain.
[90]	Blockchain- and cryptography-based	In this paper, the author stored the cryptographic hash of device metadata in a blockchain and stored the actual data in the cloud for scalability.

4.2.3. Addressing RQ-3

The third RQ for this study is "Which healthcare application achieves provenance security, confidentiality, integrity, availability, metadata protection, scalability, provenance overhead, unforgeability, and GDPR compliance?" We answer these attributes using Yes 'Y'/No 'N' and * which means partially satisfied. Also, maturity level 'ML' is considered in this question as an additional attribute. The maturity level can be architecture 'A', proposed 'P', implemented 'I', or evaluated 'E'. We summarize the results in Table 7 based on these data-provenance-related attributes. After creating Table 7 for a comparison for existing research, we discuss the attributes in the next discussion sections.

Table 7. Comparison of research articles.

Ref.	Applied Field	Technologies Used	ML	PS	C	I	A	MP	S	PO	U	GDPR
[91]	Brazilian hemotherapy	Provenance data model	I	N	Y	N	N	N	N	N	N	N
[92]	Healthcare data	HLF, BC, and smart contract	I	Y	Y	Y	Y	N	N	N	N	N
[93]	Medical data sharing	BC and smart contract	I	Y	Y	Y	Y	N	Y	Y	N	N

Table 7. *Cont.*

Ref.	Applied Field	Technologies Used	ML	PS	C	I	A	MP	S	PO	U	GDPR
[94]	Mobile health data	BC and trusted execution environment	E	Y	Y	Y	Y	Y	N	Y	Y	N
[95]	E-health system	BC	P	Y	N	Y	N	N	Y	N	N	N
[18]	Electronic medical record	Digital watermarking	I	Y	N	Y	N	N	N	N	N	N
[36]	Healthcare data	PROV and Blockchain	I	Y	Y	Y	Y	N	Y	N	Y	N
[50]	Healthcare scenario	Not mentioned	P	Y	Y	Y	Y	N	Y	Y	Y	N
[62]	Controlled medication	Ethereum BC, smart contract, and IPFS	I	Y	Y	Y	Y	Y	Y	Y	N	N
[66]	Big data	Digital signature and encryption	P	Y	N	N	N	N	Y	Y	N	N
[96]	Healthcare	Blockchain	P	Y	Y	Y	Y	Y	Y	N	N	N
[7]	IoT healthcare application	-	A	Y	N	N	N	N	N	N	N	N
[80]	Health monitoring system	BC and smart contract	E	Y	Y	Y	Y	N	Y	Y	Y	N
[82]	Health supply chain	Blockchain, Ethereum, smart contract	I	N	N	Y	Y	N	*	N	N	N
[84]	Healthcare data	Semantic web technologies	I	Y	Y	Y	Y	N	N	N	N	N
[85]	Healthcare data	Semantic web technologies	I	Y	Y	Y	N	N	N	N	N	N
[97]	Mobile-based health data	WebView, XCode	E	Y	N	Y	N		N	Y	N	N
[98]	E-health	Blockchain	I	Y	Y	Y	Y	N	N	N	Y	N
[99]	E-health data	Not mentioned	I	N	N	N	N	N	N	N	N	N
[100]	Healthcare	Blockchain, smart contract	I	N	Y	Y	Y	N	*	N	N	N
[101]	Patient data	Consortium BC	I	Y	Y	Y	Y	N	N	N	N	N
[102]	COVID vaccination	Blockchain and smart contract, IPFS	I	Y	Y	Y	Y	N	N	N	N	N
[103]	Medical data	RDBMS	P	Y	N	N	N	N	N	N	N	N

4.2.4. Addressing RQ-4

RQ-4: What are the major research gaps in the domain of data provenance in healthcare systems?

In previous questions, we discussed the technologies used for data provenance and healthcare systems/applications which achieved the data provenance. In this section, the most important question "Finding the Research gap" is addressed. Although a lot of work is done in this field, there are still research gaps that need to be filled. The most important research gaps are as follows:

(a) When the provenance is stored, the occupation of the data subject is not considered. Some jobs are highly sensitive, e.g., civil servants, government employees, armed forces, researchers who work in atomic plants, etc., and others have less sensitive information. We have found in our SLR that provenance is not classified and prioritized according to its sensitivity.

(b) Provenance metadata protection is not discussed in detail in the literature. Metadata consist of sensitive and non-sensitive attributes, and adversaries can attack using non-sensitive metadata. The classification between sensitive and non-sensitive metadata stored in provenance is very important for user privacy protection. This can be addressed by researchers in the future.

(c) The provenance storage of data subjects is directly related to privacy. There is a need to perform plenty of work in this regard. If the GDPR rules are wrapped into provenance data, privacy may be preserved. In our SLR, we have found a few research articles that emphasize the importance of the GDPR as it relates to healthcare data provenance. There is a need to address this issue not only on a conceptual level but also in an implementation context.

(d) The GDPR highlights the importance of the data subject's consent for data handling. Unfortunately, considerable work is not carried out for healthcare data provenance.

(e) Current research is achieving data provenance using blockchain technology. A public blockchain has issues with GDPR compliance and scalability. The existing web-based or cloud services-based healthcare systems deal with GDPR and scalability issues efficiently but are susceptible to a singular point of failure. The combination of web, cloud, and blockchain, either public or private-based approaches, may solve the issues of GDPR compliance and scalability.

5. Discussion

In this section, we have presented our timeline and taxonomy of data provenance in the healthcare domain to uncover and analyze them easily. We considered some important features and technologies which are important to achieve provenance and provide a comparison in Table 7. The techniques are log-based, cryptography-based, ontology-based, and blockchain-based. The essential requirements of data provenance are provenance security, compliance, storage, unforgeability, overhead, revocability, and metadata protection.

5.1. Taxonomy

Figure 10 presents the taxonomy of data provenance in healthcare. Requirements for data provenance and research work carried out are explained in the following sub-sections.

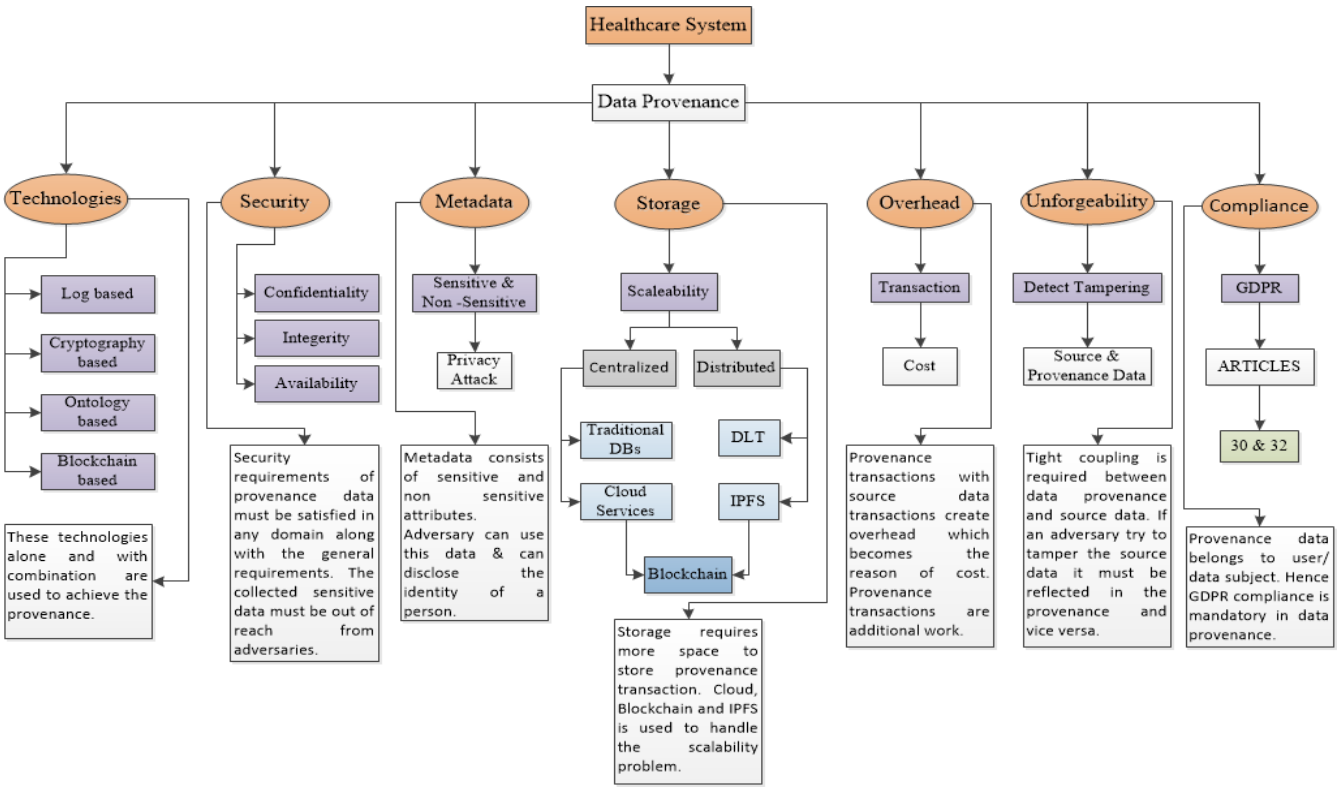

Figure 10. Taxonomy of data provenance in healthcare.

5.1.1. Technologies

During our investigation of data provenance, we identified four distinct types of technologies [34,37]. Figure 1 depicts some of the author's work using these technologies to achieve provenance. These technologies can also be used in combination to achieve data provenance, like in [36,39,62]. The authors in [14,15] proposed log-based provenance. The authors in [16,18,34] used cryptography-based solutions for achieving provenance. For ontology-based provenance, there exist some ontologies [23,31]. Data provenance in the blockchain can be achieved through blockchain transactions that record data operations [19]. One blockchain-based provenance example is provchain [20]. Some authors achieved provenance using smart contracts [21,40]. These approaches help in achieving data provenance using blockchain. We found that, in recent years, the trend is toward using blockchain technology to achieve data provenance. It provides many built-in features required for provenance and it can work well with many existing technologies.

5.1.2. Security

A data provenance system must fulfill general security requirements, e.g., confidentiality, integrity [34,50], and availability [15,34,63]. Data provenance may reveal private information. So, encryption of data provenance is very important with source data encryption. Queries of data provenance and the results of queries in response in any data provenance system must be secure and must not reveal any sensitive information. Data inside the provenance store must be immutable. We found that a blockchain-based solution for data provenance helps in maintaining the integrity of not only source data but provenance, too. In the end, the provenance of data must be available for users at any time from anywhere. In the case of patient data in an emergency case, they are subject to high availability. We found some existing research works [50,94,96,98,102] which are meeting these requirements in their work. In [7,66], the authors discussed the security of provenance but did not individually mention these requirements.

5.1.3. Storage

The authors in [64] state that provenance data can be stored in either a centralized or distributed manner, offering flexibility in the storage approach. Commonly, storage scalability is measured in terms of performance and capacity. Though blockchain technology is effective at meeting data provenance requirements, it is very hard, due to systematic design, to search and query provenance information. But when we use centralized approaches, the maintenance is very difficult to manage. Both approaches have advantages and disadvantages. The authors in [104] said that scalability refers to metadata capturing. It is very important how much metadata are captured in provenance to manage storage scalability. In the taxonomy, we highlighted in the end that mostly the authors [62,96,105] are using blockchain with cloud and IPFS to manage this issue.

5.1.4. Metadata

Provenance data, also known as lineage or pedigree, refer to a specific kind of metadata that pertain to an entity. They document the origin and utilization of a data source [33,34]. The significance of provenance metadata surpasses that of the actual data, as it can potentially be exploited by malicious actors to carry out privacy breaches. In particular, metadata that characterizes the patient as the data source requires enhanced privacy protection, especially in the context of smart health applications. The authors in [96] proposed a solution based on blockchain technology that captures and stores provenance metadata on the blockchain using smart contracts. They store the data off-chain and store the corresponding provenance on the chain. Metadata consist of sensitive and non-sensitive attributes, and the GDPR has a clear distinction between sensitive and non-sensitive personal data. Gender, date of birth, place of birth, postcode, etc. can disclose the privacy of the data subject. In the existing research trending in healthcare systems related to data provenance, we did not find special attention to the issue of sensitive and non-sensitive data in metadata. There is

a need to create such information systems in which data provenance must be stored and classified according to the sensitivity of the metadata. For example, a patient's disease history consists of highly sensitive information [65]. It is also important for the recent trend of storing provenance metadata in the blockchain.

5.1.5. Overhead

Less overhead is important for high performance in any application [66]. The impact of provenance collection must be minor [67]. In their study, the authors prioritized minimizing overhead as a design objective to enhance the performance of the proposed system [66]. The authors in [50] proposed a cloud-based data provenance system where they discussed the storage and performance overhead and how to optimize them. In a blockchain-based system, there must be low overhead for accessing and storing provenance data. This directly affects the scalability issue because data provenance transactions put an extra burden on the application with source data transactions. Blockchain-based approaches used for data provenance pay the cost of overhead due to provenance transactions. There is a need to control this cost.

5.1.6. Unforgeability

It refers to the forging of data provenance with fake data as well as with source data. Provenance must be tightly coupled with its source data, and the provenance system must detect if an adversary tries to forge fake data. Someone can forge a medical report to avoid an investigation of misdiagnosis [50]. So, the proposed system must have this feature of unforgeability. The authors in [106,107] presented protocols for forgeability in cloud-based systems. In the healthcare domain, very few authors [36,94,98] have discussed and fulfilled this requirement.

5.1.7. Compliance

Various types of information such as phone numbers, email addresses, and locations have the potential to unveil the identity of an individual. Swift advancements in technology and the widespread digitization across various sectors, including healthcare, have caused significant concerns among citizens regarding their privacy. The medical records of patients make up healthcare data [52], which is a crucial component in this context. According to GDPR Article 9, health data are a special category of data. There is a need to develop GDPR-compliant healthcare systems which also capture provenance. There is ongoing work to develop healthcare systems that are GDPR-compliant. The authors in [53,54] used a logical deletion method to solve the issue of Article 17 of the GDPR. Article 7 of the GDPR emphasizes consent management. A data subject's data cannot be stored without prior consent [55], and the authors in [56] proposed a solution according to GDPR compliance. In this SLR, one of our focuses is to find healthcare systems that are complying with GDPR in terms of data provenance. Article 30 of GDPR belongs to recording processing activities and Article 32 belongs to the storage of processing.

We found no work in which the authors discussed these articles in existing research on achieving provenance in healthcare systems. Keeping compliance in mind while developing provenance systems may open new directions.

5.2. Timeline

We have analyzed four basic technologies to achieve data provenance and listed them in the taxonomy. We also figured out the ongoing trend of the research community through observing their interests in achieving data provenance using these technologies. The timeline depicts the interest of the research community in a graphical view to more easily understand other researchers. In Figure 11, the timeline is presented from 2010 to 2022, where the technologies are mapped alongside the years. On the *X*-axis, the years are plotted, and technologies are plotted on the *Y*-axis. We created the timeline through keeping our focus on the healthcare domain. In 2006, authors argued that traditional log

recording is suited for a single-node system [30]. Though the limitations were overcome by authors in [15], the report of authors in a survey-based study presented that log-file-based provenance is not suited to the distributed nature of current systems [34]. We found a log-based approach from 2012 to 2017 for healthcare domain data provenance. Then, the trend shifts toward other techniques. In 2020, the authors [73] proposed an approach with log files with a distributed ledger to achieve data provenance. Their effort demonstrated that log-based techniques may be helpful in the future to work in combination with other technologies. Cryptography-based technologies are always in use for data provenance and work with other technologies for data provenance. This technology helps in identifying the origin of data efficiently [108]. Watermarking [18], encryption [66], IPFS [62,102], digital signatures [16,66], DHT [105], etc. are all based on cryptography. But, in the case of data processing history, this technology cannot store records [1]. From 2010 to 2022, we have found this technology helpful in data provenance. From 2018 to 2021, ontology-based provenance was used in the healthcare domain. Capturing provenance using ontology is very flexible and can cover as many domains as possible [31,109]. It is also very helpful in provenance analysis [110]. Several ontologies exist in the literature for data provenance. Ontologies store the provenance record [111] and there is always a need to apply a technique on ontology to protect it from malicious entities. From all these technologies we found that the adoption of blockchain technology is growing day by day due to its features for achieving data provenance. From 2016 to 2022, the authors adopted blockchain technology. In research question 3, we listed several papers in which authors proposed and implemented blockchain-based solutions in combination with IPFS [62,102], smart contracts [83,101], and the PROV model [36]. This technology is very useful in distributed-nature application scenarios [32,33]. It is very complex in terms of implementation context [1]. Also, scalability is a challenge for this technology. In the taxonomy, we have highlighted the authors [62,96,105] who are using blockchain with cloud and IPFS to manage this issue.

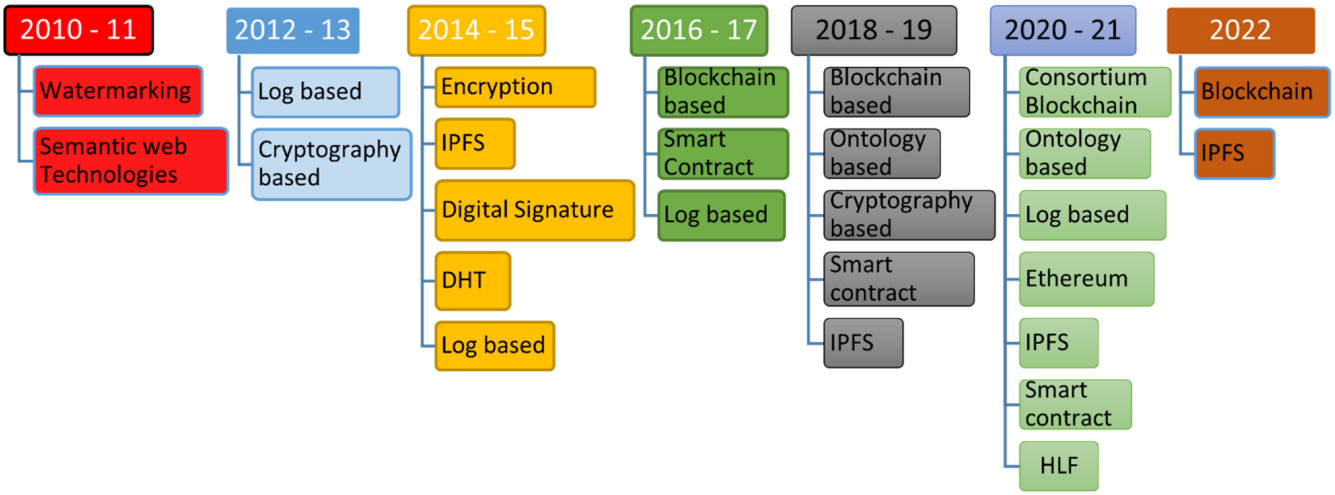

Figure 11. Timeline of technologies used for data provenance in healthcare.

6. Contributions

In this SLR, we have explored the research work performed for achieving provenance in the healthcare domain. We have discussed the technologies used to achieve provenance in prior literature. We have also identified the research gaps based on the SLR, and based on our findings we have also made several suggestions for future research directions. Figure 12 presents the open research directions. The implications are as follows.

1. This paper presents the techniques used to achieve data provenance in healthcare. We found four techniques, i.e., log-based, cryptography-based, blockchain-based, and ontology-based techniques for achieving data provenance, as discussed in Table 5. Further, we explored the work carried out to achieve the data provenance with the

combination of these techniques, as discussed in Table 6. We found that limited work was performed to achieve provenance using the combination of these technologies. Researchers can find possible solutions using these techniques, as discussed in Table 6 or can introduce new techniques to achieve data provenance.

2. This SLR discussed the GDPR compliance issue in the field of healthcare data provenance, which was not discussed in any other SLR. In Table 7, we have explored different attributes. GDPR compliance is one of them, and the purpose was to explore compliance issues because provenance data belongs to entities, i.e., the data subject who is a patient in case of healthcare data, or the data controller, etc. GDPR Article 30 is related to the record of processing activities. In data provenance, we store actions performed by the data subject or data controller to achieve trustworthiness. Though the authors, in their research work, achieved data provenance, they did not discuss the GDPR compliance issues. Very rare work is done in this regard. Dealing with the GDPR compliance issues when achieving data provenance in healthcare will help and set a new direction for research in the future.
3. In existing healthcare systems, authors have achieved data provenance. But they did not explain which types of data they are storing during the operations or actions performed by the user or administrator. Because provenance consists of sensitive and non-sensitive information, an adversary can attack and may reveal the identity of the user. So, when you are collecting provenance, it is important to classify the metadata into different categories and protect them from adversaries using encryption techniques. Provenance security and provenance metadata are very important attributes from the research point of view in the future.
4. Blockchain technology ensures the immutability of data. But when dealing with big data, especially in the medical field, it becomes challenging. As the data grow in the blockchain, the performance of the application degrades. Every new miner will download a complete copy of the data. Also, the GDPR conflicts with blockchain for storing data. The data stored on a blockchain cannot be easily modified or deleted; it can create compliance issues under the GDPR. For big data applications in the future, we recommend using cloud services with blockchain technology. Through integrating blockchain and cloud services, organizations can leverage the strengths of both technologies to achieve a more comprehensive solution. Blockchain will deal with security and the cost of maintenance, and the cloud environment will deal with storage issues, performance, and compliance. Its decentralized and tamper-resistant nature enhances data security, making it difficult for unauthorized individuals to manipulate or tamper with the stored information.
5. The timeline in Figure 11 shows that the trend in achieving data provenance after 2015 is by means of blockchain technology. Blockchain works using transactions. The transactions become the reason for communication overhead in the blockchain as the network grows. The provenance may put an additional burden on the blockchain due to additional transactions of data provenance in terms of cost. The evaluation of data provenance transactions with data storage is also required.
6. Another important direction that is already in use for capturing provenance is the use of ontologies. Capturing provenance using ontology is very flexible and can cover as many domains as possible. Several ontologies exist in the literature for data provenance. It is also very helpful in provenance visualization and analysis. Ontologies store the provenance record, and researchers can apply encryption techniques to provenance data in an ontology to protect it from malicious entities. In ontologies, one can classify the data in different classes according to the importance of data.

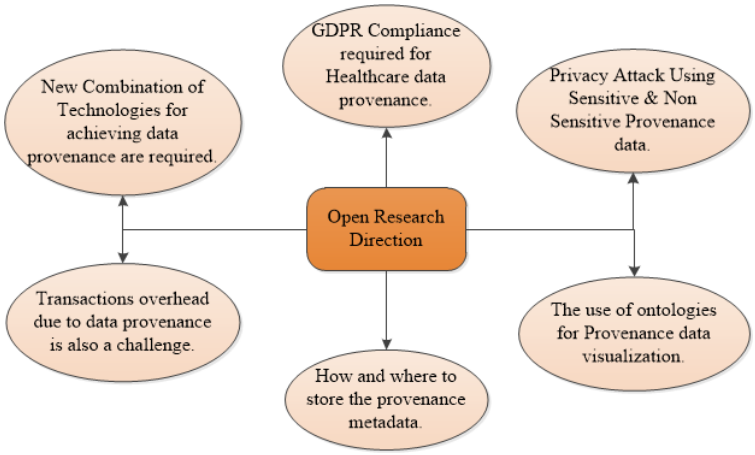

Figure 12. Open research directions.

Limitations

Despite the various benefits of conducting a systematic analysis, certain drawbacks need to be considered, e.g., bias in the collection, publication, the extraction of data, and misclassification [30]. During the SLR, we found many domains in which data provenance exists. The focus of this SLR was purely limited to research articles that only addressed data provenance for healthcare data and GDPR compliance. We have developed an efficient research protocol to achieve the maximum possible number of papers. Inclusion and exclusion criteria were set to ensure that the focus was on the research topic.

7. Conclusions and Future Work

The SLR answers the questions set in the research methodology of the current state of the art in data provenance in the healthcare domain. We have discussed the types of data provenance and the technologies used to achieve it in the healthcare domain. We have identified blockchain technology as the latest technology in healthcare systems to achieve immutable data provenance. Our findings show that several challenges still require more research, i.e., provenance security, provenance communication overhead, scalability, revocation, and GDPR compliance. Very little effort is made on GDPR compliance data provenance in the literature.

In the future, this SLR may assist as a reference in this field. Potential researchers, with the aid of our contributions and identification of research gaps, may design a new model or architecture. The use of blockchain technology with existing data-provenance-based healthcare systems in particular may overcome the existing problems with more exciting solutions.

Author Contributions: Conceptualization, M.A., A.R.D. and M.H.; methodology, M.A., A.R.D. and A.K.; validation, M.A., M.H., J.K. and A.K.; formal analysis, M.H., A.R.D. and J.K.; investigation, M.A., A.R.D. and M.H.; resources, M.H., A.K. and J.K.; data curation, M.A., J.K., A.K.; writing—original draft, M.A. and A.R.D.; preparation, M.A., A.R.D. and M.H.; writing—review and editing, M.A., A.R.D., J.K. and A.K.; supervision, M.H. and J.K.; project administration, A.K. and J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea under Grant NRF-2022R1A2C1012037.

Acknowledgments: The authors are thankful to Science Foundation Ireland (Nos. [13/RC/2106_P2] and [20/SP/8955]) at the ADAPT SFI Research Centre at Maynooth University. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme.

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

References

1. Buneman, P.; Khanna, S.; Tan, W.C. Data provenance: Some basic issues. In Proceedings of the International Conference on Foundations of Software Technology and Theoretical Computer Science, New Delhi, India, 13–15 December 2020; Springer: Berlin/Heidelberg, Germany, 2000.
2. Ma, T.; Wang, H.; Cao, Z.; Yong, J.; Zhao, Y. Access Control Management with Provenance in Healthcare Environments. In Proceedings of the 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanchang, China, 4–6 May 2016.
3. Sabaa, H.; Panda, B. Data Authentication and Provenance Management. In Proceedings of the 2007 2nd International Conference on Digital Information Management, Lyon, France, 28–31 October 2007; Volume 1.
4. Yan, Z.; Yu, X.; Ding, W. Context-aware verifiable cloud computing. *IEEE Access* **2017**, *5*, 2211–2227. [CrossRef]
5. Yu, X.; Yan, Z.; Vasilakos, A.V. A survey of verifiable computation. *Mob. Netw. Appl.* **2017**, *22*, 438–453. [CrossRef]
6. Yu, X.; Yan, Z.; Zhang, R. Verifiable outsourced computation over encrypted data. *Inf. Sci.* **2019**, *479*, 372–385. [CrossRef]
7. Elkhodr, M.; Alsinglawi, B.; Alshehri, M. Data Provenance in the Internet of Things. In Proceedings of the 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), Krakow, Poland, 16–18 May 2018.
8. Jagadish, H.V.; Olken, F. Database management for life sciences research. *ACM SIGMOD Rec.* **2004**, *33*, 15–20. [CrossRef]
9. Hossain, M.M.; Hasan, R.; Zawoad, S. Trust-IoV: A Trustworthy Forensic Investigation Framework for the Internet of Vehicles (IoV). In Proceedings of the 2017 IEEE International Congress on Internet of Things (ICIOT), Honolulu, HI, USA, 25–30 June 2017.
10. Sigwart, M.; Borkowski, M.; Peise, M.; Schulte, S.; Tai, S. Blockchain-Based Data Provenance for the Internet of Things. In Proceedings of the 9th International Conference on the Internet of Things, Bilbao, Spain, 22–25 October 2019.
11. Hossain, M.M.; Hasan, R.; Zawoad, S. Data Provenance Trusted Model in Cloud Computing. In Proceedings of the 2013 International Conference on Research and Innovation in Information Systems (ICRIIS), Kuala Lumpur, Malaysia, 27–28 November 2013.
12. Hajnal, Á.; Kifor, T.; Pedone, G.; Varga, L.Z. Benefits of provenance in home care. *Stud. Health Technol. Inform.* **2007**, *126*, 330.
13. Moreau, L.; Groth, P.; Miles, S.; Vazquez-Salceda, J.; Ibbotson, J.; Jiang, S.; Munroe, S.; Rana, O.; Schreiber, A.; Tan, V.; et al. The provenance of electronic data. *Commun. ACM* **2008**, *51*, 52–58. [CrossRef]
14. Jiang, W.; Hu, C.; Pasupathy, S.; Kanevsky, A.; Li, Z.; Zhou, Y. Understanding customer problem troubleshooting from storage system logs. In Proceedings of the 7th Conference on File and Storage Technologies, San Francisco, CA, USA, 24 February 2009; pp. 43–56.
15. Suen, C.H.; Ko, R.K.L.; Tan, Y.S.; Jagadpramana, P.; Lee, B.S. S2logger: End-to-end Data Tracking Mechanism for Cloud Data Provenance. In Proceedings of the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, IUCC-2013, Melbourne, Australia, 16–18 July 2013.
16. Priyadharshini, M.D.; Ananth, C. A secure hash message authentication code to avoid certificate revocation list checking in vehicular adhoc networks. *Int. J. Appl. Eng. Res.* **2015**, *10*, 1250–1254. [CrossRef]
17. Sultana, S.; Shehab, M.; Bertino, E. Secure provenance transmission for streaming data. *IEEE Trans. Knowl. Data Eng.* **2012**, *25*, 1890–1903. [CrossRef]
18. Tharaud, J.; Wohlgemuth, S.; Echizen, I.; Sonehara, N.; Müller, G.; Lafourcade, P. Privacy by Data Provenance with Digital Watermarking—A Proof-of-Concept Implementation for Medical Services with Electronic Health Records. In Proceedings of the 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, Germany, 15–17 October 2010; pp. 510–513. [CrossRef]
19. Salman, T.; Zolanvari, M.; Erbad, A.; Jain, R.; Samaka, M. Security services using blockchains: A state of the art survey. *IEEE Commun. Surv. Tutor.* **2018**, *21*, 858–880. [CrossRef]
20. Liang, X.; Shetty, S.; Tosh, D.; Kamhoua, C.; Kwiat, K.; Njilla, L. Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In Proceedings of the 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), Madrid, Spain, 14–17 May 2017; pp. 468–477.
21. Neisse, R.; Steri, G.; Nai-Fovino, I. A Blockchain-Based Approach for Data Accountability and Provenance Tracking. In Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy, 29 August–1 September 2017.
22. Guarino, N. (Ed.) *Formal Ontology in Information Systems: Proceedings of the First International Conference (FOIS'98), 6–8 June, Trento, Italy*; IOS Press: Amsterdam, The Netherlands, 1998; Volume 46.
23. Dalpra, H.L.; Costa, G.C.B.; Sirqueira, T.F.M.; Braga, R.M.; Campos, F.; Werner, C.M.L.; David, J.M.N. Using Ontology and Data Provenance to Improve Software Processes. In Proceedings of the Brazilian Seminar on Ontologies, São Paulo, Brazil, 8–11 September 2015.
24. Hasan, S.S.; Sultan, N.H.; Barbhuiya, F.A. Cloud Data Provenance Using IPFS and Blockchain Technology. In Proceedings of the Seventh International Workshop on Security in Cloud Computing, Auckland, New Zealand, 8 July 2019.
25. Benet, J. Ipfs-content addressed, versioned, p2p file system. *arXiv* **2014**, arXiv:1407.3561.
26. Bose, R.J.C.; Phokela, K.K.; Kaulgud, V.; Podder, S. Blinker: A Blockchain-Enabled Framework for Software Provenance. In Proceedings of the 2019 26th Asia-Pacific Software Engineering Conference (APSEC), Putrajaya, Malaysia, 2–5 December 2019; pp. 1–8.
27. Khatal, S.; Rane, J.; Patel, D.; Patel, P.; Busnel, Y. 2021. Fileshare: A Blockchain and IPFS Framework for Secure File Sharing and Data Provenance. In *Advances in Machine Learning and Computational Intelligence*; Springer: Singapore, 2021; pp. 825–833.

28. Lovell, M.; Foy, M.A. General data protection regulation May 2018 (GDPR). *Bone Jt.* **2018**, *7*, 4142. [CrossRef]
29. Krystlik, J. With GDPR, preparation is everything. *Comput. Fraud Secur.* **2017**, *2017*, 58. [CrossRef]
30. Muniswamy-Reddy, K.K.; Holland, D.A.; Braun, U.; Seltzer, M.I. Provenance-Aware Storage Systems. In Proceedings of the Usenix Annual Technical Conference, General Track, Boston, MA, USA, 30 May–3 June 2006.
31. Ram, S.; Liu, J. A New Perspective on Semantics of Data Provenance. In Proceedings of the First International Workshop on the role of Semantic Web in Provenance Management, Washington, DC, USA, 25 October 2009; Volume 526.
32. Buneman, P.; Khanna, S.; Wang-Chiew, T. Why and Where: A Characterization of Data Provenance. In Proceedings of the International Conference on Database Theory, London, UK, 4–6 January 2001; Springer: Berlin/Heidelberg, Germany, 2001.
33. Simmhan, Y.L.; Plale, B.; Gannon, D. A survey of data provenance in e-science. *ACM SIGMOD Rec.* **2005**, *34*, 31–36. [CrossRef]
34. Hu, R.; Yan, Z.; Ding, W.; Yang, L.T. A survey on data provenance in IoT. *World Wide Web* **2020**, *23*, 1441–1463. [CrossRef]
35. Gao, M.; Jin, C.Q.; Wang, X.L.; Tian, X.X.; Zhou, A.Y. A survey on management of data provenance. *Chin. J. Comput.* **2010**, *33*, 373–389. [CrossRef]
36. Massi, M.; Miladi, A.; Margheri, A.; Sassone, V.; Rosenzweig, J. *Using PROV and Blockchain to Achieve Health Data Provenance*; University of Southampton Institutional Repository: Southampton, UK, 2018; pp. 1–24.
37. Liu, J. *W7 Model of Provenance, and Its Use in the Context of Wikipedia*; The University of Arizona: Tucson, AZ, USA, 2011.
38. Yazici, I.M.; Aktas, M.S. A novel visualization approach for data provenance. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6523. [CrossRef]
39. Kim, H.M.; Laskowski, M. Toward an ontology-driven blockchain design for supply-chain provenance. *Intell. Syst. Account. Financ. Manag.* **2018**, *25*, 18–27. [CrossRef]
40. Ramachandran, A.; Kantarcioglu, D.M. Using blockchain and smart contracts for secure data provenance management. *arXiv* **2017**, arXiv:1709.10000.
41. Kalogriopoulos, N.A.; Baran, J.; Nimunkar, A.J.; Webster, J.G. Electronic Medical Record Systems for Developing Countries. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MI, USA, 3–6 September 2009.
42. Onik, M.M.H.; Aich, S.; Yang, J.; Kim, C.S.; Kim, H.C. Blockchain in Healthcare: Challenges and Solutions. In *Big Data Analytics for Intelligent Healthcare Management*; Academic Press: Cambridge, MA, USA, 2019; pp. 197–226.
43. Banerjee, S.; Hemphill, T.; Longstreet, P. Wearable devices and healthcare: Data sharing and privacy. *Inf. Soc.* **2018**, *34*, 49–57. [CrossRef]
44. Manyazewal, T. Using the World Health Organization health system building blocks through survey of healthcare professionals to determine the performance of public healthcare facilities. *Arch. Public Health* **2017**, *75*, 50. [CrossRef]
45. World Health Organization. *Everybody's Business—Strengthening Health Systems to Improve Health Outcomes: WHO's Framework for Action*; World Health Organization: Geneva, Switzerland, 2007.
46. Kifor, T.; Varga, L.Z.; Álvarez, S.; Vázquez-Salceda, J.; Willmott, S. Privacy Issues of Provenance in Electronic Healthcare Record Systems. In Proceedings of the 1st International Workshop on Privacy and Security in Agent-Based Collaborative Environments (PSACE 2006), Hakodate, Japan, 9 May 2006.
47. Pittman, P. Health Services Research in 2020: Data and methods needs for the future. *Health Serv. Res.* **2010**, *45*, 1431. [CrossRef] [PubMed]
48. Lopez, M.H.; Holve, E.; Sarkar, I.N.; Segal, C. Building the informatics infrastructure for comparative effectiveness research (CER): A review of the literature. *Med. Care* **2012**, *50*, S38–S48. [CrossRef] [PubMed]
49. Groth, P.; Jiang, S.; Miles, S.; Munroe, S.; Tan, V.; Tsasakou, S.; Moreau, L. *An Architecture for Provenance Systems*; University of Southampton: Southampton, UK, 2006.
50. Asghar, M.R.; Ion, M.; Russello, G.; Crispo, B. Securing Data Provenance in the Cloud. In Proceedings of the International Workshop on Open Problems in Network Security, Lucerne, Switzerland, 9 June 2011; Springer: Berlin/Heidelberg, Germany, 2011.
51. Mourby, M.; Mackey, E.; Elliot, M.; Gowans, H.; Wallace, S.E.; Bell, J.; Smith, H.; Aidinlis, S.; Kaye, J. Are 'pseudonymised'data always personal data? Implications of the GDPR for administrative data research in the UK. *Comput. Law Secur. Rev.* **2018**, *34*, 222233. [CrossRef]
52. Koutli, M.; Theologou, N.; Tryferidis, A.; Tzovaras, D.; Kagkini, A.; Zandes, D.; Karkaletsis, K.; Kaggelides, K.; Miralles, J.A.; Oravec, V.; et al. Secure IoT e-Health Applications Using VICINITY Framework and GDPR Guidelines. In Proceedings of the 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, 29–31 May 2019; p. 263270.
53. Truong, N.B.; Sun, K.; Lee, G.M.; Guo, Y. GDPR-compliant personal data management: A blockchain-based solution. *IEEE Trans. Inf. Secur.* **2020**, *15*, 17461761. [CrossRef]
54. Ma, S.; Guo, C.; Wang, H.; Xiao, H.; Xu, B.; Dai, H.-N.; Cheng, S.; Yi, R.; Wang, T. Nudging Data Privacy Management of Open Banking Based on Blockchain. In Proceedings of the 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN), Yichang, China, 16–18 October 2018; p. 7279. [CrossRef]
55. Ahmed, J.; Yildirim, S.; Nowostaki, M.; Ramachandra, R.; Elezaj, O.; Abomohara, M. GDPR Compliant Consent Driven Data Protection in Online Social Networks: A Blockchain-Based Approach. In Proceedings of the 3rd International Conference on Information and Computer Technologies (ICICT), San Jose, CA, USA, 9–12 March 2020; p. 307312. [CrossRef]

56. Saglam, R.B.; Aslan, C.B.; Li, S.; Dickson, L.; Pogrebna, G. A Data-Driven Analysis of Blockchain Systems' Public Online Communications on GDPR. In Proceedings of the IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS), Oxford, UK, 3–6 August 2020; p. 2231. [CrossRef]
57. Kitchenham, B.; Brereton, O.P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15. [CrossRef]
58. Khan, M.M.; Ibrahim, R.; Ghani, I. Cross domain recommender systems: A systematic literature review. *ACM Comput. Surv.* **2017**, *50*, 1–34. [CrossRef]
59. Yumna, H.; Khan, M.M.; Ikram, M.; Ilyas, S. Use of Blockchain in Education: A Systematic Literature Review. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Yogyakarta, Indonesia, 8–11 April 2019; Springer: Cham, Switzerland, 2019; pp. 191–202.
60. Aly, M.; Khomh, F.; Haoues, M.; Quintero, A.; Yacout, S. Enforcing Security in Internet of Things Frameworks: A Systematic Literature Review. *Internet Things* **2019**, *6*, 100050. [CrossRef]
61. Alomar, N.; Alsaleh, M.; Alarifi, A. Social authentication applications, attacks, defense strategies and future research directions: A systematic review. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1080–1111. [CrossRef]
62. Musamih, A.; Jayaraman, R.; Salah, K.; Hasan, H.R.; Yaqoob, I.; Al-Hammadi, Y. Blockchain-Based Solution for the Administration of Controlled Medication. *IEEE Access* **2021**, *9*, 145397–145414. [CrossRef]
63. Rabah, K. Challenges & opportunities for blockchain powered healthcare systems: A review. *Mara Res. J. Med. Health Sci.* **2017**, *1*, 45–52.
64. Dogan, G. A Survey of Provenance in Wireless Sensor Networks. *Adhoc Sens. Wirel. Netw.* **2016**, *30*, 21–45.
65. Can, O.; Yilmazer, D. A Privacy-Aware Semantic Model for Provenance Management. In Proceedings of the Research Conference on Metadata and Semantics Research, Karlsruhe, Germany, 27–29 November 2014; Springer: Cham, Switzerland, 2014.
66. Imran, A.; Agrawal, R.; Walker, J.; Gomes, A. A Layer Based Architecture for Provenance in Big Data. In Proceedings of the 2014 IEEE International Conference on Big Data, Big Data 2014, Washington, DC, USA, 27–30 October 2014; pp. 29–31. [CrossRef]
67. Cheah, Y.W.; Canon, R.; Plale, B.; Ramakrishnan, L. Milieu: Lightweight and Configurable Big Data Provenance for Science. In Proceedings of the 2013 IEEE International Congress on Big Data, Santa Clara, CA, USA, 27 June–2 July 2013; pp. 46–53.
68. Ghoshal, D.; Plale, B. Provenance from Log Files: A BigData Problem. In Proceedings of the Joint EDBT/ICDT 2013 Workshops, Genoa, Italy, 22 March 2013.
69. Ko, R.K.; Will, M.A. Progger: An Efficient, Tamper-Evident Kernel-Space Logger for Cloud Data Provenance Tracking. In Proceedings of the 2014 IEEE 7th International Conference on Cloud Computing, London, UK, 8–11 December 2014.
70. Tan, Y.S. Reconstructing Data Provenance from Log Files. Ph.D. Thesis, The University of Waikato, Hamilton, New Zealand, 2017.
71. Groth, P.; Gil, Y.; Magliacane, S. Automatic Metadata Annotation through Reconstructing Provenance. In Proceedings of the Semantic Web in Provenance Managment Workshop, Heraklion, Greece, 28 May 2012.
72. Kaaniche, N.; Belguith, S.; Laurent, M.; Gehani, A.; Russello, G. Prov-Trust: Towards a Trustworthy SGX-Based Data Provenance System. In Proceedings of the 17th International Joint Conference on e-Business and Telecommunications-Volume 3: SECRYPT, Paris, France, 8–10 July 2020; ScitePress: Setúbal, Portugal, 2020.
73. Ma, S.; Zhang, X.; Xu, D. Protracer: Towards Practical Provenance Tracing by Alternating between Logging and Tainting. In Proceedings of the Network and Distributed System Security Symposium, San Diego, CA, USA, 21–24 February 2016; Volume 2.
74. Vikhe, A.; Desai, P. Data provenance verification for secure hosts using advance cryptography algorithm. *Int. J. Comput. Appl.* **2014**, *88*, 25–27. [CrossRef]
75. Sultana, S.; Ghinita, G.; Bertino, E.; Shehab, M. A Lightweight Secure Provenance Scheme for Wireless Sensor Networks. In Proceedings of the 2012 IEEE 18th International Conference on Parallel and Distributed Systems, Singapore, 17–19 December 2012; pp. 101–108.
76. Wang, C.; Bertino, E. Sensor network provenance compression using dynamic bayesian networks. *ACM Trans. Sens. Netw. (TOSN)* **2017**, *13*, 1–32. [CrossRef]
77. Siddiqui, M.S.; Rahman, A.; Nadeem, A. Secure data provenance in IoT network using bloom filters. *Procedia Comput. Sci.* **2019**, *163*, 190–197. [CrossRef]
78. Yang, Y.; Liu, X.; Guo, W.; Zheng, X.; Dong, C.; Liu, Z. Multimedia access control with secure provenance in fog-cloud computing networks. *Multimed. Tools Appl.* **2020**, *79*, 10701–10716. [CrossRef]
79. Javaid, U.; Aman, M.N.; Sikdar, B. Blockpro: Blockchain Based Data Provenance and Integrity for Secure Iot Environments. In Proceedings of the 1st Workshop on Blockchain-enabled Networked Sensor Systems, Shenzhen, China, 4 November 2018.
80. Sigwart, M.; Borkowski, M.; Peise, M.; Schulte, S.; Tai, S. A secure and extensible blockchain-based data provenance framework for the internet of things. *Pers. Ubiquitous Comput.* **2020**, 1–15. [CrossRef]
81. Shetty, S.; Red, V.; Kamhoua, C.; Kwiat, K.; Njilla, L. Data provenance assurance in the cloud using blockchain. In Proceedings of the Disruptive Technologies in Sensors and Sensor Systems, Anaheim, CA, USA, 2 May 2017; Volume 10206, pp. 125–135.
82. Musamih, A.; Salah, K.; Jayaraman, R.; Arshad, J.; Debe, M.; Al-Hammadi, Y.; Ellahham, S. A blockchain-based approach for drug traceability in healthcare supply chain. *IEEE Access* **2021**, *9*, 9728–9743. [CrossRef]
83. Gong, J.; Lin, S.; Li, J. Research on Personal Health Data Provenance and Right Confirmation with Smart Contract. In Proceedings of the 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 20–22 December 2019; Volume 1.

84. Can, O.; Yilmazer, D. Improving privacy in health care with an ontology-based provenance management system. *Expert Syst.* **2020**, *37*, e12427. [CrossRef]
85. Cadenhead, T. *Secured Data Provenance Using Semantic Web Technologies*; The University of Texas at Dallas: Richardson, TX, USA, 2011.
86. Hartig, O.; Zhao, J. Publishing and consuming provenance metadata on the web of linked data. In Proceedings of the Provenance and Annotation of Data and Processes: Third International Provenance and Annotation Workshop, IPAW 2010, Troy, NY, USA, 15–16 June 2010.
87. Sahoo, S.S.; Sheth, A.P. Provenir Ontology: Towards a Framework for Escience Provenance Management. 2009. Available online: https://corescholar.libraries.wright.edu/knoesis/80 (accessed on 8 May 2022).
88. Kuo, T.T.; Kim, H.E.; Ohno-Machado, L. Blockchain distributed ledger technologies for biomedical and health care applications. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 1211–1220. [CrossRef]
89. Rahman, M.S.; Khalil, I.; Mahawaga Arachchige, P.C.; Bouras, A.; Yi, X. A novel architecture for tamper proof electronic health record management system using blockchain wrapper. In Proceedings of the 2019 ACM International Symposium on Blockchain and Secure Critical Infrastructure, Auckland, New Zealand, 8 July 2019; pp. 97–105.
90. Ali, S.; Wang, G.; Bhuiyan, Z.A.; Jiang, H. Secure Data Provenance in Cloud-Centric Internet of Things via Blockchain Smart Contracts. In Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, China, 8–12 October 2018; pp. 991–998.
91. Sembay, M.J.; de Macedo, D.D.J.; Dutra, M.L. A Proposed Approach for Provenance Data Gathering. *Mob. Netw. Appl.* **2021**, *26*, 304–318. [CrossRef]
92. Margheri, A.; Masi, M.; Miladi, A.; Sassone, V.; Rosenzweig, J. Decentralised provenance for healthcare data. *International J. Med. Inform.* **2020**, *141*, 104197. [CrossRef] [PubMed]
93. Xia, Q.I.; Sifah, E.B.; Asamoah, K.O.; Gao, J.; Du, X.; Guizani, M. MeDShare: Trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access* **2017**, *5*, 14757–14767. [CrossRef]
94. Hardin, T.; Kotz, D. Amanuensis: Information provenance for health-data systems. *Inf. Process. Manag.* **2021**, *58*, 102460. [CrossRef]
95. Kubendiran, M.; Singh, S.; Sangaiah, A.K. Enhanced security framework for e-health systems using blockchain. *J. Inf. Process. Syst.* **2019**, *15*, 239–250.
96. Bawany, N.Z.; Qamar, T.; Tariq, H.; Adnan, S. Integrating Healthcare Services Using Blockchain-Based Telehealth Framework. *IEEE Access* **2022**, *10*, 36505–36517. [CrossRef]
97. Lomotey, R.K.; Deters, R. Mobile-Based Medical Data Accessibility in mHealth. In Proceedings of the 2014 2nd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, Oxford, UK, 8–11 April 2014; pp. 91–100. [CrossRef]
98. Li, S.; Zhang, Y.; Xu, C.; Cheng, N.; Liu, Z.; Shen, X.S. Besure: Blockchain-based cloud-assisted ehealth system with secure data provenance. In Proceedings of the 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS), Tokyo, Japan, 25–28 June 2021; pp. 1–6.
99. Sun, Y.; Lu, T.; Gu, N. A Method of Electronic Health Data Quality Assessment: Enabling Data Provenance. In Proceedings of the 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD), Wellington, New Zealand, 26–28 April 2017; pp. 233–238. [CrossRef]
100. Zhuang, Y.; Sheets, L.R.; Chen, Y.-W.; Shae, Z.-Y.; Tsai, J.J.P.; Shyu, C.-R. A Patient-Centric Health Information Exchange Framework Using Blockchain Technology. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2169–2176. [CrossRef]
101. Sabir, A.; Fetais, N. A Practical Universal Consortium Blockchain Paradigm for Patient Data Portability on the Cloud Utilizing Delegated Identity Management. In Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2–5 February 2020; pp. 484–489. [CrossRef]
102. Odoom, J.; Soglo, R.S.; Danso, S.A.; Xiaofang, H. A Privacy-Preserving Covid-19 Updatable Test Result and Vaccination Provenance Based on Blockchain and Smart Contract. In Proceedings of the 2019 International Conference on Mechatronics, Remote Sensing, Information Systems and Industrial Information Technologies (ICMRSISIIT), Accra, Ghana, 20–22 December 2019; pp. 1–6.
103. Ma, T.; Yong, J.; Wang, H.; Zhao, Y. Causal Dependencies of Provenance Data in Healthcare Environment. In Proceedings of the 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Calabria, Italy, 6–8 May 2015; pp. 643–648. [CrossRef]
104. Holland, D.A.; Braun, U.; Maclean, D.; Muniswamy-Reddy, K.-K.; Seltzer, M.I. Choosing A Data Model and Query Language for Provenance. In Proceedings of the 2nd International Provenance and Annotation Workshop (IPAW'08), Salt Lake City, UT, USA, 17–18 June 2008; Springer: Berlin/Heidelberg, Germany, 2008.
105. Seneviratne, O.; Kagal, L. Enabling Privacy through Transparency. In Proceedings of the 2014 Twelfth Annual International Conference on Privacy, Security and Trust, Toronto, ON, Canada, 23–24 July 2014; pp. 121–128. [CrossRef]
106. Muniswamy-Reddy, K.-K.; Macko, P.; Seltzer, M.I. Provenance for the Cloud. In Proceedings of the USENIX Conference on File and Storage Technologies (FAST '10), San Jose, CA, USA, 23–26 February 2010; Volume 10.
107. Muniswamy-Reddy, K.-K.; Seltzer, M. Provenance as first class cloud data. *ACM SIGOPS Oper. Syst. Rev.* **2010**, *43*, 11–16. [CrossRef]

108. Gibb, G.; Zeng, H.; McKeown, N. Outsourcing Network Functionality. In Proceedings of the First Workshop on Hot Topics in Software Defined Networks, Helsinki, Finland, 13 August 2012.
109. Groth, P.; Moreau, L. PROV-Overview. An Overview of the PROV Family of Documents. 2013. Available online: https://eprints.soton.ac.uk/356854/ (accessed on 13 May 2022).
110. Oliveira, W.; Ambrósio, L.M.; Braga, R.; Ströele, V.; David, J.M.; Campos, F. A framework for provenance analysis and visualization. *Procedia Comput. Sci.* **2017**, *108*, 1592–1601. [CrossRef]
111. Pandit, H.J.; Lewis, D. Modelling Provenance for GDPR Compliance using Linked Open Data Vocabularies. In Proceedings of the 5th Society, Privacy and the Semantic Web—Policy and Technology Workshop (PrivOn 2017), Co-Located with ISWC 2017, Vienna, Austria, 22 October 2017.