WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Information Traceability in a Heterogeneous Environment: Challenges and Solutions

Séraphin Buambaka Buambaka

UPKAN, RDC

SUMMARY

Information traceability is a fundamental pillar to ensure data integrity, security and compliance within organizations. It not only tracks the data journey through the different processes, but also provides the transparency necessary to maintain stakeholder trust. Its tools, adapted to the specific needs of each sector, play an essential role in the implementation of this traceability, thus ensuring effective data management and compliance with current regulations. Aeronautical administration in particular benefits greatly from these systems. These help maintain the high quality and safety standards required in this area. Information traceability is a key component that supports strategic and operational decision-making to protect the company's information assets.

Keywords: **Informational traceability, heterogeneous environment, heterogeneous data, Standardization, Virtualization, Mutualization.**

1. INTRODUCTION

Information traceability is an essential concept in a world where data flows between diverse systems, heterogeneous platforms and diverse environments. Whether in administration, health, logistics, finance or scientific research, the ability to track the origin, transformations and interactions of information is crucial.

Traceability is the ability to trace the history, use or location of an entity through recorded identifications[1].

Information traceability consists in monitoring the origin, transformations and interactions of data through various systems, it is a major issue in a heterogeneous environment. Key issues facing us include:

- How to ensure consistency and reliability of information despite diversity?
- Why align information from different systems?
- What can be done to manage this metadata effectively?

Data consistency and reliability are essential to ensure information reliability, even in a diverse environment. For this, it is necessary to take into account the regular cleaning of the data, to know their source, their update log and their integration.

With regard to the alignment of information from disparate sources, consideration should be given to the integration of information systems, their information governance and their alignment model.

To manage this data effectively, we need to apply the rules and processes to access, share, link, integrate, maintain and analyze metadata across the enterprise

2. STATE OF THE ART

1. **Ministry of Culture and Strategic Communication,** traceability of digital data: What models for data provenance?. This article discusses the importance of data provenance and traceability in the context of the semantic web. The W3 consortium has developed a model, PROV, generic, extensible and interoperable, for the origin metadata or to know the origin allows to judge its quality, to know if it meets the technical and legal requirements of a given profession [2]

[1] S. TAMAYO GIRALDO., Exploitation of traceability information for the optimization of choices in production and logistics, doctoral thesis in computer science, Université Paul Verlaine, Metz, 2011, p.18

[2] Ministry of Culture and Communication Strategy, Digital Data Traceability: What Models for Data Provenance, article published in December 2015 in France

2. **Philippe KISLIN,** TRACEABILITY, ANNOTATION AND HISTORISATION Three postures to establish trust in the collaboration of the watchdog and the decision-maker. In his article the author proposes several recommendations among which He highlights the importance of traceability, annotation and historization to establish trust in the collaboration of the watchman and the decision maker. He mentions that these aspects concern the objectivity of the traces, the completeness of the follow-up and the control of the volumes in the face of the continuity of the flow of decision and the difficulty of implementing procedures of capture and explanation of knowledge [3]

3. **Simon Tamayo Giraldo**, Exploitation of traceability information for optimizing choices in production and logistics. The author proposes the diagnosis of a production by the exploitation of traceability information, and presents the improvement actions to be implemented, He proposes tools to obtain these additional benefits. These analysis, processing and optimization tools must be adaptable to the configurations and constraints of companies and sufficiently ergonomic and robust to process traceability information flows. To perform criticality diagnosis based on traceability information, it uses an artificial neural network so that criticality diagnosis can exploit most traceability databases in production. [4]

4. **Dupuy, C.,** Analysis and design of tools for the traceability of agri-food products in order to optimize the dispersion of manufacturing batches. In his thesis, the author proposed methods and tools to optimize the traceability of agri-food products and the dispersion of manufacturing batches. In particular, he proposed models of the material flow and its monitoring, as well as the traceability registration process. In addition, he presented a mathematical model to optimize traceability in a specific case: three-level bills of materials with assembly and disassembly. [5]

5. **Jennifer LAZZERI** & Nathalie FABBE-COSTES the total traceability of supply chains: concept and theoretical model of implementation. The authors propose a theoretical analysis of supply chain traceability, in an intra- and inter-organizational vision. They stress the importance of full traceability, which is difficult to implement. Faced with the lack of empirical work and the absence of a clear theoretical framework, they propose a construction of the concept of total traceability and develop a theoretical model on the conditions of implementation of such a system. They then discuss the contributions and limitations of this model[6].

6. **E. HADJIPAVLOU.,** BigData, Surveillance and Trust: The issue of traceability in the airport environment. The author makes a significant contribution to understanding the implications of Big Data and digital traceability in the specific context of the airport environment. It critically and comprehensively examines the notion of digital traces in the era of Big Data and the relationship between the notions of monitoring and trust. Big Data refers to the massive production of data that represents a valuable wealth of benefits. [7]

Regarding our study, for better information traceability in a heterogeneous environment, we propose several solutions to address these challenges. Information traceability in a heterogeneous environment is a complex but essential challenge. There are many issues, ranging from data security to data accessibility and standardization. The combination of the cloud **who solution** that **combines Traceability,** consolidation**e and** pooling, as well as the application of common standards, are all tracks to meet these challenges.

To complete our study, we used the following methods:

- Comparative method: Identify different contexts or situations finally retain differences and similarities;
- Descriptive method: This method consists of mapping data on a specific aspect. It is often quantitative and based on previous studies;
- Field research: which has helped us observe data directly in the field in order to understand real practices and challenges encountered in a heterogeneous environment.

3. CHALLENGES OF TRACEABILITY IN A HETEROGENEOUS ENVIRONMENT

Data traceability plays a crucial role in information management within companies. It tracks the origin, validity and integrity of data throughout its lifecycle: from creation to deletion. In a heterogeneous environment, there are several challenges:

1. Diversity of Sources

In a heterogeneous environment, data comes from a variety of sources: sensors, databases, applications, third-party systems, etc. Each source has its own format, semantics, and rules. Traceability becomes complex when this data needs to be aggregated, compared or analyzed.

2. Standardization and Harmonization

[3] P, KISLIN., Traceability, Annotation and Historization: Three postures to establish trust in the collaboration of the watchman and the decision-maker, Editions Lavoisier, Dans le cahier du numérique, April 2009
[4] S, TOMAYO-GIRALDO., Exploitation of traceability information for the optimization of choices in production and logistics, doctoral thesis in computer science and automation, Paul Verlaine University of Metz, France, 2011
[5] C, DUPUY., Analysis and design of tools for the traceability of agri-food products in order to optimize the dispersion of manufacturing batches, Ph.D. thesis. National Institute of Applied Sciences of Lyon, 2004
[6] J, LAZZERI and N, FABRE-COSTES., Total traceability of supply chains: concept and theoretical model of implementation, article published in the Revue Française de Gestion Industrielle Vol. 33, N° 1 / 2014, pp.55–79
[7] E. HADJIPAVLOU., Big Data, Surveillance and Trust: The question of traceability in the airport environment, PhD thesis in computer science, Université Côte d'Azur, 2016

Data standardization is a major issue. How to align information from different systems? What can I do to manage units of measurement, codes, identifiers? Standardization avoids misinterpretation and facilitates traceability.

3. Metadata and Documentation

Metadata is "data beyond data". They provide contextual information: origin, date of creation, author, etc. Documenting this metadata is essential to understand the context of the information and ensure its traceability.

4. Unique Identifiers

The use of unique identifiers is fundamental. Each piece of information must have an identifier that follows it throughout its life cycle. This facilitates research, updating and traceability.

5. Security and Privacy

Traceability raises questions of security and confidentiality. How can we ensure that only authorised persons have access to the information? What can be done to protect sensitive data while ensuring traceability?

4. IMPORTANCE AND ROLES OF TRACEABILITY IN DATA CONSOLIDATION

Traceability in data consolidation is fundamental for several reasons[8]:

- ✓ Ensure data quality: traceability makes it possible to track the origin, validity and integrity of data. It also ensures that the data is accurate and up to date.
- ✓ Meet compliance standards: Many laws and regulations require companies to ensure the traceability of their data, including privacy, data security and governance.
- ✓ Identify and resolve problems: Data traceability helps to trace errors and data problems, and quickly identify the sources of these difficulties in order to resolve them quickly.
- ✓ Facilitate collaboration: traceability facilitates collaboration by allowing team members to easily understand the history and origin of data. This also facilitates communication and problem solving.
- ✓ Improve decision-making: By ensuring data traceability, organizations can have accurate and reliable information to make informed, data-driven decisions.

Data traceability plays a crucial role in information management within companies[9]:

- ✓ Data Manager: is responsible for the overall management of the organization's data, including data traceability. It is responsible for defining policies and procedures to ensure the traceability of data, as well as overseeing its collection, analysis and management.
- ✓ Compliance Officer: is responsible for compliance with data management laws and regulations. He or she must ensure that the organization complies with legal requirements for data traceability, including privacy, data security and data governance.
- ✓ Data Security Officer: is responsible for the security of the organization's data. In order to ensure their traceability, it must ensure that they are stored and managed securely, using appropriate security techniques to ensure data confidentiality, integrity and availability.
- ✓ Data Quality Manager: is responsible for the quality of the organization's data. It must ensure that data is accurate, complete and reliable, using data quality management techniques to ensure its quality and traceability throughout its life cycle[10].
- ✓ Data Analyst: is responsible for collecting, analyzing and interpreting data. They must ensure that they are properly labelled, documented and tracked throughout their life cycle to ensure traceability.
- ✓ Data Users: is responsible for using data to make informed decisions. He must ensure that the data he uses is accurate, reliable and traceable, in order to guarantee the quality of his decisions. In case of failure, it must inform the data/IT teams of the corrective actions to be taken.

5. How to ensure traceability?

Data traceability involves continuous and rigorous work. It requires a combination of appropriate practices and policies that will anchor effective processes in the organization. For this, it is necessary to: [11]

[8] J-L, VIRUERA., Traceability, Tools, Methods and Practices, Organization Editions, Paris 2005, p.49
[9] https://datavalue-consulting.com/tracabilite-des-donnee, accessed on Monday 08 April 2024 at 10:00'
[10] P, KISLIN., Traceability, Annotation and Historization Three postures to establish trust in the collaboration of the watchman and the decision-maker, Editions Lavoisier, in the digital notebook, April 2009, p.40
[11] https://datavalue-consulting.com/tracabilite-des-donnee, accessed Monday 08 April 2024 at 12h00'

- ✓ Identify and label data: each dataset must be clearly identified and tagged with metadata describing its content, format, source and owner.
- ✓ Track changes to data: Every time data is changed, changes must be recorded and documented to ensure traceability.
- ✓ Record data access: Every time a person or system accesses data, this must be recorded and documented to ensure traceability.
- ✓ Use data management tools: data management tools, such as database management systems (DBMS), can help ensure data traceability by automating metadata collection and management.
- ✓ Establish data security policies: Clear and well-defined policies can help ensure data traceability by establishing procedures for their management and protection.
- ✓ Conduct regular audits and reviews: regular audits and reviews of data can help identify errors and traceability issues, and correct them quickly.

VI. Tools used for data traceability

There are several tools used for data traceability, which help track the origin, transformations and end use of data. Some commonly used tools include:

- ✓ Metadata management tools: They document the origin, transformations and destination of data.
- ✓ Database Management Systems (DBMS): They offer features such as logging, transaction logs and audit features.
- ✓ Data Quality Management Tools: These tools monitor, clean and standardize data to ensure data quality.
- ✓ Advanced analysis tools: they optimize logistics operations using the data collected to identify trends and evaluate performance.[12]

These tools are essential to ensure good data governance and to help organizations meet compliance standards, improve decision-making and facilitate collaboration between teams.

The implementation of traceability can be broken down into 4 main steps:

- ✓ Identify the data elements;
- ✓ follow the origin: follow the listed items and identify the origin of each of them
- ✓ List sources and links: create a spreadsheet or table to reference sources and link elements that can be combined;
- ✓ create a map: create maps for each device and a main map to group all the main elements.

Choosing the best data traceability tool depends on the specific needs, the size of the business, and the types of data you manage. However, there are well-established tools known for their efficiency and reliability, for example Talend, which is often cited for its ability to manage and create mapping.

In aeronautical administration, the implementation of a traceability tool must meet strict safety and quality standards. For example, solutions such as FabTracer interface with the information system and enable it to be populated with relevant and secure field data.

VII. SOLUTIONS FOR EFFECTIVE TRACEABILITY

To achieve this informational traceability, we must follow these guidelines:

1. Data Management Systems

Implement centralised data management systems. Data warehouses, databases and document management systems enable information to be stored, organised and managed in a coherent way.

2. Monitoring data flows

Monitor data flows between systems. Identify breakpoints, transformations and any loss of information. Regular audits guarantee data quality and compliance with traceability rules.

3. Interdisciplinary collaboration

Traceability is a multidisciplinary challenge. Involve experts in IT, data management, security and the business. Together, they can design solutions tailored to the environment.

4. Training and awareness

Make users aware of the challenges of traceability. Train them in best practices, tools and methodologies. Traceability is everyone's business.

[12] J-L, VIRUERA.,op.cit, p.59

Information traceability is a complex field that encompasses the ability to track the origin, transformations and interactions of data across a variety of systems. While there is no single algorithm that solves all aspects of traceability, here are some of the approaches and techniques used in this context:

1. Data Flow Tracking Algorithms:

- ✓ These algorithms make it possible to track the path of data through different systems. They record processing steps, transformations, and waypoints.
- ✓ For example, in supply chains, parcel tracking algorithms are used to trace the movement of goods from shipment to delivery.

2. Normalization and Harmonization Algorithms:

- ✓ To ensure traceability, it is essential to align data from heterogeneous sources. Algorithms are used to standardize formats, units of measurement, and terminologies.
- ✓ Data harmonization facilitates comparison and aggregation.

3. Metadata Management Algorithms:

- ✓ Metadata provides contextual information about the data. Algorithms are used to extract, store, and manage this information.
- ✓ For example, metadata can indicate where the data came from, how good it is, when it was created, etc.

4. Security and Privacy Algorithms:

- ✓ Traceability must comply with security and confidentiality rules. Encryption, restricted access, and rights management algorithms are used to protect data while allowing traceability.

5. Search and Correlation Algorithms:

- ✓ When it comes to tracing information, search and correlation algorithms are used. They make it possible to find links between different data and to reconstruct events.
- ✓ For example, in criminal investigations, algorithms can link clues from various sources to solve a case.

Information traceability is based on a combination of specific algorithms adapted to the context and the data concerned. Each field (aeronautics, logistics, healthcare, finance, etc.) may require different approaches for effective traceability.

For better information traceability in a heterogeneous environment, we want the combination of the cloud solution that combines traceability, consolidation and pooling

1**. The cloud** provides a secure framework in a heterogeneous environment, thus contributing to the traceability and protection of critical information.

Data consolidation: Without mandatory consolidation, there is a risk that indicators will appear in company publications that give a truncated and biased view of the group's performance.[13] Consolidation is key to effectively managing information in a world where data comes from multiple sources and formats. It provides a unified and consistent view, supporting informed decision-making and operational performance. Data consolidation is very important for:

- ✓ Unification of data sources
- ✓ Process Optimization;
- ✓ Data Reliability
- ✓ Informed decision-making ;

Three approaches are used to consolidate data:

- ETL (Extract, Transform and Load)

- Data virtualization

- Data warehousing

1. Multi-tenancy: Unify disparate sources into a cohesive view, making it easier to access data. It also optimizes the use of IT resources, avoids redundancies, and reduces costs by sharing infrastructure and services. Despite resource sharing, each client or system remains isolated at the logical level. Data is secured and segregated to ensure confidentiality and integrity.

In addition, the use of web applications can greatly facilitate traceability with the following ease:

[13] F. DEPOERS., The consolidation of environmental data: Issues and Practices, éditions HAL open sciences, Nice, 2010, p.4

In addition, the use of web applications can greatly facilitate traceability with the following ease:

- Recording the history of each interaction with users. This practice tracks user actions, pages visited, queries made, etc.;

- Use of internal web applications: these applications record information about users, transactions and data changes;

- Traceability of web flows: web technologies record traces at each interaction. Internet flows through the web are recorded in databases, thus making it possible to trace users' activities;

- Interoperability with the existing information system: traceability is ensured by the exchange of data between these systems.

Indeed, web applications play a very crucial role in information traceability, by recording the interactions, flows and actions of users in a heterogeneous environment.

Cloud and web application security are key concerns for protecting data and systems in a heterogeneous environment

A. For the Cloud: Control and manage permissions to access cloud resources, Protect data in transit and at rest, Encrypt data.

Example of an RSA Algorithm to Encrypt and Decrypt the Message

```
ALGORITHM Generate KeysRSA
INPUT: None
OUTPUT: clé_publique (n, e), clé_privée (d, n)
FUNCTION CalculateMMI(e, m)
RETURN D
END FUNCTION
EncryptRSA(message, clé_publique)
RETURN message_chiffré
END FUNCTION
DECRYPT FUNCTIONRa(message_chiffré, clé_privée)
   Uses the RSA algorithm to decrypt the message encrypted with the private key
   Example: message_déchiffré = message_chiffré^d % n
   (Actual implementation uses operations on large numbers)
    ...
   RETURN message_déchiffré
END FUNCTION
MAIN FUNCTION
   Key Generation
   p, q = GeneratePrimeNumbers()
   n = p * q
   m = (p - 1) * (q - 1)
   e = ChooseSuperscript(m) // Example: e = 65537 (current value)
   d = ComputeMMI(e, m)
   clé_publique = (n, e)
   clé_privée = (d, n)
   Example of use
   message = "Confidential data"
   message_chiffré = EncryptRSA(message, clé_publique)
```

```
    message_déchiffré = DecipherRSA(message_chiffré, clé_privée)
    SHOW "Encrypted Message: " + message_chiffré
    SHOW "Decrypted Message: " + message_déchiffré
END FUNCTION
```

B. For Web Apps: Protect apps from attacks, enforce authentication, verify and filter user input.

Example of an AES Symmetric Encryption Algorithm

```
AES Encryption ALGORITHM
INPUT: Data (string), key (string)
OUTPUT: donnée_chiffrée (string)
FUNCTION EncryptAES(data, key)
    donnée_chiffrée = AES(data, key)
    donnée_chiffrée = AES(data, key)
    RETURN donnée_chiffrée
END FUNCTION
DECRYPT FUNCTIONAES(donnée_chiffrée, key)
    donnée_déchiffrée = AES(donnée_chiffrée, key)
    RETURN donnée_déchiffrée
END FUNCTION
MAIN FUNCTION
    data = "Sensitive data to be encrypted"
    key = "MacléBuasera85"
    donnée_chiffrée = EncryptAES(data, key)
    donnée_déchiffrée = DecryptAES(donnée_chiffrée, key)
        SHOW "Numerical data: " + donnée_chiffrée
    SHOW "Decrypted data: " + donnée_déchiffrée
END FUNCTION
```

CONCLUSION

Information traceability in a heterogeneous environment is a complex but essential challenge. The issues are numerous, ranging from data security to data accessibility and standardization. The proposed solutions, such as the use of traceability tools, the integration of the cloud for consolidation and pooling, as well as the application of common standards, are all avenues to meet these challenges.

Consolidation techniques like ETL, virtualization, and data warehousing play a crucial role in managing heterogeneous data. The importance of cloud and web application security was emphasized, with the use of encryption algorithms such as AES for web applications and RSA for the cloud. These measures help to ensure the integrity and confidentiality of the data, while allowing its traceability.

It is clear that information traceability is an essential component of data management in a heterogeneous environment. Despite the challenges, the potential benefits in terms of operational efficiency, transparency and accountability are considerable. It is therefore imperative for organizations to continue to explore and implement innovative solutions for information traceability. This will not only allow them to protect their information assets, but also allow them to leverage them to gain a competitive advantage.

In short, information traceability, despite its challenges, remains an essential component of an organization's digital strategy. Efforts in this area will undoubtedly be beneficial for the future of data management.

BIBLIOGRAPHY

1. C, DUPUY., Analysis and design of tools for the traceability of agri-food products in order to optimize the dispersion of manufacturing batches, Ph.D. thesis. National Institute of Applied Sciences of Lyon, 2004

2. E. HADJIPAVLOU., Big Data, Surveillance and Trust: The question of traceability in the airport environment, PhD thesis in computer science, Université Côte d'Azur, 2016

3. F. DEPOERS., The consolidation of environmental data: Issues and Practices, HAL open sciences editions, Nice, 2010

4. https://datavalue-consulting.com/tracabilite-des-donnee, accessed on Monday 08 April 2024 at 10:00'

5. https://datavalue-consulting.com/tracabilite-des-donnee, accessed Monday 08 April 2024 at 12:00'

6. J, LAZZERI and N, FABRE-COSTES., Total Traceability of Supply Chains: Concept and Theoretical Model of Implementation, article published in the Revue française of Industrial Management Vol. 33, N° 1 / 2014.

7. J-L, VIRUERA., Traceability, Tools, Methods and Practices, Organization Editions, Paris, 2005

8. Ministry of Culture and Communication Strategy, Digital Data Traceability: Which Models for Data Provenance? Article published in December 2015 in France

9. P, KISLIN., Traceability, Annotation and Historization: Three Postures to Establish Trust in the Collaboration of the Watchman and the Decision-Maker, Editions Lavoisier, In digital Notebook, April 2009

10. S, TOMAYO-GIRALDO., Exploitation of traceability information for the optimization of choices in production and logistics, doctoral thesis in computer science and automation, Paul Verlaine University of Metz, France, 2011