International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 1, Jan-Feb 2024

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Implementing Formalised Provenances to Enhance Administrative Ontologies

Oluseyi Ayodeji Oyedeji[1], Ibiyinka Temilola Ayorinde[2]
[1]Department of Computer Science, University Of Ibadan, Ibadan, Nigeria
Email: seyioyedeji31@gmail.com
[2]Department of Computer Science, University Of Ibadan, Ibadan, Nigeria
Email:temiayorinde@yahoo.com

----------**********----------

Abstract:

The need for clarity of processes and interoperability of administrative applications has led to the creation of ontologies about the administrative domain. However, there is a gap in terms of traceability of the concepts. Unlike other domains, governance and administration is more trusted, usable, and effective when the entities and activities can be authenticated and verified. This research introduces a formalised provenance approach to ascertain the origin of concepts such as process status, digital resources, and personnel within the administrative domain. Firstly, an existing administrative ontology is identified as a case study. Provenance points are extracted from the ontology such as process provenance, document provenance and personnel provenance. The knowledge for verifying these provenance points is outlined in plain language and then formalised using description logic. The description logics are then formatted using Web ontology language (owl). Finally, Competency questions are used to validate the provenance ontology using SPARQL query. This research is an extension of existing administrative ontologies by adding provenance, thereby increasing the level of quality and trust in administrative applications when they make use of the ontology. It also improves the level of data integrity when these administrative applications share data between each other.

Keywords **—Ontology, Provenance, Administration, Digital governance, Knowledge representation**

----------**********----------

I. INTRODUCTION

Knowledge representation has been used to formalise real world information such that computer-based systems, both simple and even complex-robotic systems [8] can access this information, reason, and even make conclusions. This is why knowledge representation is an important aspect of artificial intelligence [1] and it has been applied to various domains ranging from health, engineering, transportation to governance and so many other areas. In knowledge representation, the formal specification and representation of terms within a domain is commonly called ontology. Ontology enables explicit definition of terms, reusability, sharing amongst others. Administration or governance is one of such research areas where Ontology has

been fairly applied over the years. Administration is a very sensitive aspect of every organisation, in fact, it is the heart of every human-based system because it exists in government, companies, large corporations and even start-ups. Although, administration in every organisation is not the same due to difference in work culture or internal policies, however, it is very necessary that systems that are built to either automate or support the processes conform to a standard framework as much as possible. Ontology has levels and it separates data from knowledge structures, for example, an organisation might have a Chief Executive Officer (CEO) while another has a Director, however both the CEO and Director can be referred to as "Head" in a broader perspective. This is what ontology does, it creates a standardised way of formalizing domain knowledge [19].

Administration or governance ontologies that exist focus on important terms and relationships between the concepts. The missing parts are the origin of the concepts that were formalised. In other areas such as supply chain [20] and alert systems, there has always been the need to verify the origin of concepts because it is not always about the data retrieved, it is always important to verify the source of such data. In other words, governance will also be well improved and trusted if the source of entities is verified. This research aims to implement provenances to trace the origin of administrative concepts within an existing administrative ontology. The provenances are built from scratch but attached to an existing ontology as an extension. Validation is done using competency questions in the form of queries.

II. LITERATURE REVIEW

Attempts have been made in previous research to implement provenances into ontologies. In the field of digital evolution, an ontology named OntoAvida was created to help researchers have access into the vocabulary and gain better insight into evolutionary processes and mechanism of digital organisms [15]. Besides the basic term specification and inference feature provided by OntoAvida, it also possesses the provenance ability needed to trace the source and integrity of data stored.

A systematic review was carried out by [16] to research on the effect of data provenance in healthcare and how it relates to General Data Protection Regulation (GDPR). Technologies that can help to achieve provenance were identified, hence, Ontology was detected to be an effective way to implement provenance in the review [16]. In neuroscience, a Neurobridge ontology was created to improve the findability of important data about the source and origin of neuroimaging data such as important information about study participants. The provenance ontology helped to annotate full length articles and increase the ability to find provenance terms [9].

In omics study, Gene ontology (GO) is used for identification of underlying functional patterns. However, due to the evolving nature of biological knowledge, there is need for representation of GO version, hence the need for provenance to trace the version and related parameters for better interpretation and reproducibility A model that complies with FAIR principle and based on provenance ontology (PROV-O) was used to address the lack of provenance problem [11]. Another approach improved upon the Gene Ontology by making use of PPIntergrator which contains a module to describe provenance using triples [18].

An ontology design pattern that models scientific taxonomy focused on modelling provenance of concepts. This adds more context to the domain and improves governance. It also adds tag to provenance and other metadata information to bridge gaps between different ontologies [6]. To build trustworthy AI systems, it is very important to have an inclination of the origin of data when using knowledge graphs to integrate data from different sources. An ontology for tracking provenance and context of computational observations was created [12].

Another area where provenance has proven very useful and effective is in alert detection in intrusion detection systems due to the need to understand the origin of attacks [7], this reduced the number of false alerts enormously. Other areas where provenance have been integrated with ontology include Electronic Laboratory Notebooks (ELN) [14], ProvVact for tracking vaccination history [13], SAO ontology for tracing system accountability and audit [5], Internet of things (IOT) to trace the origins of medical data [10], ProvAnalyser for validating scientific workflow and to determine the quality of generated data products [4], METACLIP for tracing and understanding provenance of climate data products [2] , OntoPedigree for capturing traceability information of Pedigree within the supply chain [17].

Development of the provenance begins with the identification of an existing administrative ontology as seen in Figure 1. The administrative ontology chosen as a case for this study is an ontology for postgraduate administration which uses The Postgraduate college, University of Ibadan as a reference point [21]. The ontology consists of very important concepts which includes 347 classes and more than 1800 axioms. This makes the ontology very suitable and appropriate for this study.

Provenance points are identified from the chosen ontology. These are the points on the ontology where provenances will be added, they are the concepts that need tracing and tracking for effectiveness and robustness of the ontology. They are Process Provenance, Document Provenance and Personnel Provenance.

III. METHODOLOGY

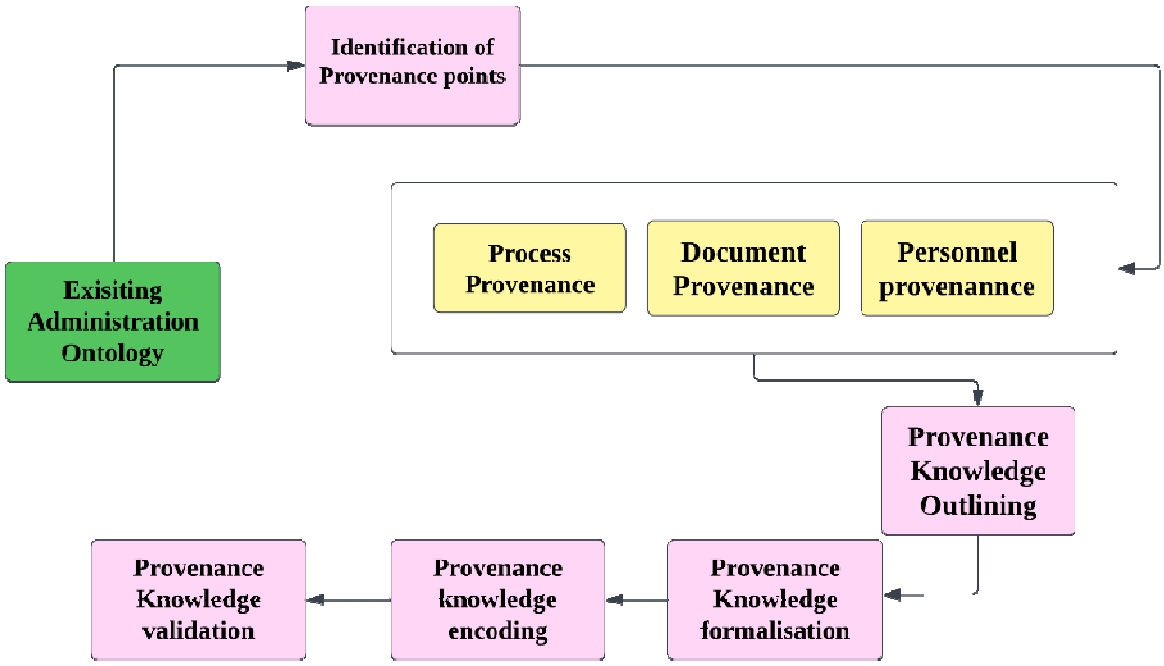

Figure 1: Provenance development methodology

A. *Process provenance*

Processes within the administrative ontology need provenance. There is a need to understand the root of the processes and how they move along the administrative pathway. Usually, these are the main building blocks of an administration even in non-digital mediums. As seen in figure 2, the elements of Process provenance are:

a. Process status
b. Event history
c. Process attribute

The process status refers to the stage at which a process is at a given time. For example, an application process could have a status of initiated, in-progress, submitted etc. Event history refers to the specific events that cause change in status of the processes. The process attribute refers to properties of the process.

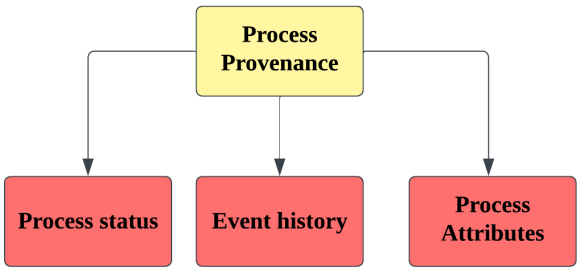

Figure 2: Process Provenance

B. *Document provenance*

Resources like documents used in administration need provenance. There are sensitive and non-sensitive documents involved in administration and regardless of the type of document, there is need to verify that the document originated from within the same administrative niche and to also see the movement along the administrative path. The elements involved in the Document provenance as seen in figure 3 include:

a. Authentication status
b. Location history
    i. Source history
    ii. Destination history
c. Document attributes

Authentication status is needed to verify the origin of the document while the location history helps to keep track of the places the document has been and passed through within the administration path. The document attributes basically refer to Document properties.

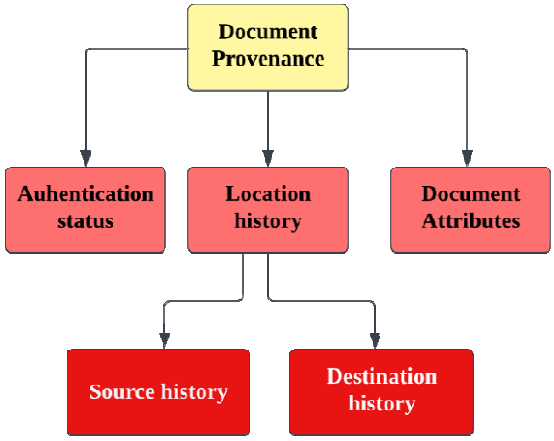

Figure 3: Document Provenance

C. *Personnel Provenance*

Personnels within the administrative ontology also needs provenance. These are the actors who carry out activities and act upon documents. The elements as seen in figure 4 are:

a. Designation authentication
b. Employment authentication

c. Personnel attribute

Designation authentication helps to prove that personnel truly possess the title they claim while employment authentication proves that the personnel originated from the organisation at some point. The personnel attributes are basic personnel properties.

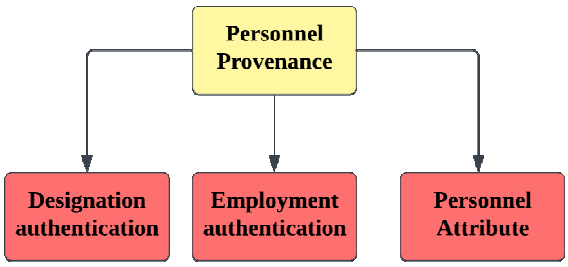

Figure 4: Personnel Provenance

After identification of the provenance points, Knowledge relating to the elements of the provenance points are outlined. Examples of knowledge at each provenance point is outlined as follows:

Process. e.g an application has a status of under-review based on:

- ✓ Proof 1: Application form has been purchased by prospective postgraduate student.
- ✓ Proof 2: Prospective student fills application form
- ✓ Proof 3: Prospective students have submitted application forms.
- ✓ Proof 4: Postgraduate college has received an application form.
- ✓ Proof 5: Postgraduate college is reviewing applications.

Documente.g. A digital asset or document moves from a state of in-authentication to a state of authentication based on:

- ✓ Proof 1: The digital asset has been signed by an authority.
- ✓ Proof 2: The singer is the real authority.

Personnele.g.A person is considered an examinations officer based on:

- ✓ Proof 1: The person is a valid staff.
- ✓ Proof 2: The valid staff belongs to the examination section.
- ✓ Proof 3: The valid staff has the authority of the Postgraduate college to be the examinations officer.

Knowledge formalisation is done using description logics. The general template for the provenance defined in this work is given as:

$$\boldsymbol{C}(\Delta\, \boldsymbol{A}\colon \boldsymbol{X},\boldsymbol{Y}) \rightarrow ((\boldsymbol{P})\boldsymbol{o} \ldots\ldots (\boldsymbol{P})\boldsymbol{n}) \quad \text{..... Equation (1)}$$

Where C = Concept in context

A = Attribute of concept

$\Delta\, \boldsymbol{A}$ = Change in attribute or transition in attribute

X = Old state of attribute

Y = New state of attribute

P = Proof of action (Axiom)

n = Number of proofs

Each of the outlined knowledge is translated into description logic. An example of the specific description logics for an application status is:

$$\boldsymbol{InitiatedApplication}(\Delta\, state\colon none - existing, initiated) \rightarrow$$

$$(\boldsymbol{p1}\colon Person \sqsubseteq isActorIn.APEV08.PurchasingApplicationForm)$$

After the description logic formalisation, the formalised knowledge is encoded into web ontology language (owl). The encoding is done

using Protégé software which has a friendly GUI instead of encoding from scratch. Finally, the ontology is tested using competency questions. Two competency questions are converted to SPARQL to test each Provenance points as seen in table 1.

TABLE 1

COMPETENCY QUESTIONS FOR VALIDATING

s/n	Competency Question	Provenance point
CQ1	What is the status of process PR23111	Process provenance
CQ2	Retrieve the event history of process PR23111	Process provenance
CQ3	Retrieve the location history of document 213313	Document provenance
CQ4	Retrieve the authentication status of document 22378	Document provenance
CQ5	Retrieve the designation authentication of personnel with name 'John Doe'	Personnel provenance
CQ6	Retrieve the employment authentication of personnel with name 'Jane Anna'	Personnel provenance

IV. IMPLEMENTATION

The Provenance ontology was implemented within the Protégé ontology editing software. The new classes, object and data properties are attached to the existing ontology. The three provenance points as implemented in Protégé interface are seen in figures 5, 8 and 11. Evaluation of the ontology is done using competency questions. The competency questions are converted into SPARQL queries and are used to query the ontology, the results are represented in figures 6, 7, 9, 10, 12 and 13 spanning the different provenance points.

In addition, ELK reasoner which is a plugin to protégé was used in checking the correctness of the ontology. This is to maintain that the ontology is consistent especially because it was extended and to ascertain that there are no contradictions.

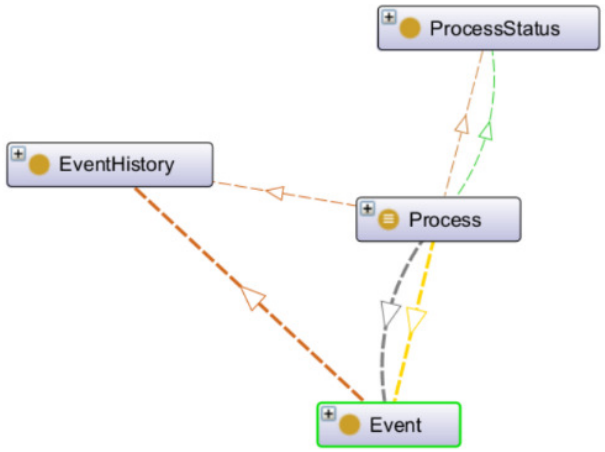

Figure 5: Process provenance view

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX io: <http://www.semanticweb.org/s361984/ontologies/2022/2/postgraduate-administration#>
SELECT ?Process ?ProcessStatus
WHERE
{
        ?Process io:identity "PR23111".
        ?Process io:hasEventStatus ?ProcessStatus.
}
```

Process	ProcessStatus
Process012	Awaiting_Approval

Figure 6: Competency question 1 – Retrieving process status.

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX io: <http://www.semanticweb.org/s361984/ontologies/2022/2/postgraduate-administration#>
SELECT ?Process ?identity ?EventHistory ?date
WHERE
{
        ?Process io:hasEventHistory ?EventHistory.
        ?Process io:identity ?identity.
        ?Process io:date ?date
}
```

Process	identity	EventHistory	date
Process012	"PR23111"	Request_made_by_Staff_0233	"21st December 2023"
Process012	"PR23111"	Request_Checked_by_Head_of_Department	"21st December 2023"
Process012	"PR23111"	Request_Signed	"21st December 2023"
Process012	"PR23111"	Request_Forwarded_to_Procurement_Department	"21st December 2023"

Figure 7: Competency question 2 – Retrieving process history.

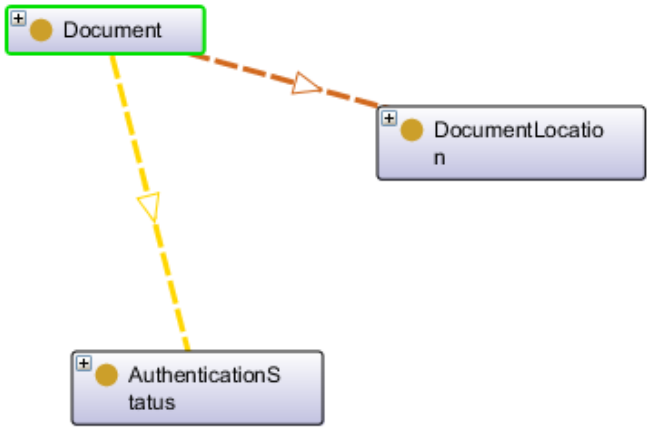

Figure 8: Document Provenance view

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX io: <http://www.semanticweb.org/s361984/ontologies/2022/2/postgraduate-administration#>
SELECT ?Document ?DocumentLocation ?date
WHERE
{
        ?Document io:hasLocation ?DocumentLocation.
        ?DocumentLocation io:date ?date
}
```

Document	DocumentLocation	date
Document_213313	FInance_Department	"21st August 2023"
Document_213313	Procurement_Department	"30th December 2023"
Document_213313	Reception	"July 7th 2023"
Document_213313	IT_Department	"11th September 2023"

Figure 9: Competency question 3 – Retrieving document trail.

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX io: <http://www.semanticweb.org/s361984/ontologies/2022/2/postgraduate-administration#>
SELECT ?Document ?identity ?authentication_status ?date
WHERE
{
        ?Document io:identity "22378".
        ?Document io:identity ?identity.
        ?Document io:authentication_status ?authentication_status.
        ?Document io:date ?date
}
```

Document	identity	authentication_status	date
Document_22378	"22378"	"Fully authenticated"	"12th Decemeber 2023"

Figure 10: Competency question 4 – Querying document authenticity

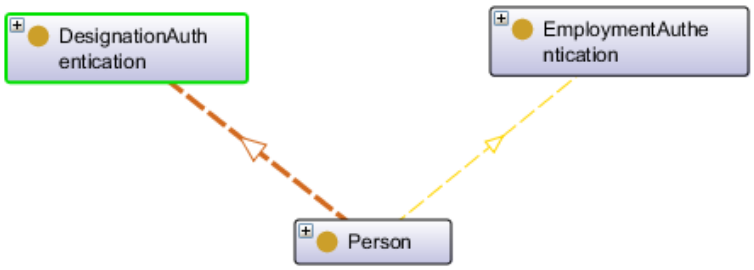

Figure 11: Personnel provenance view

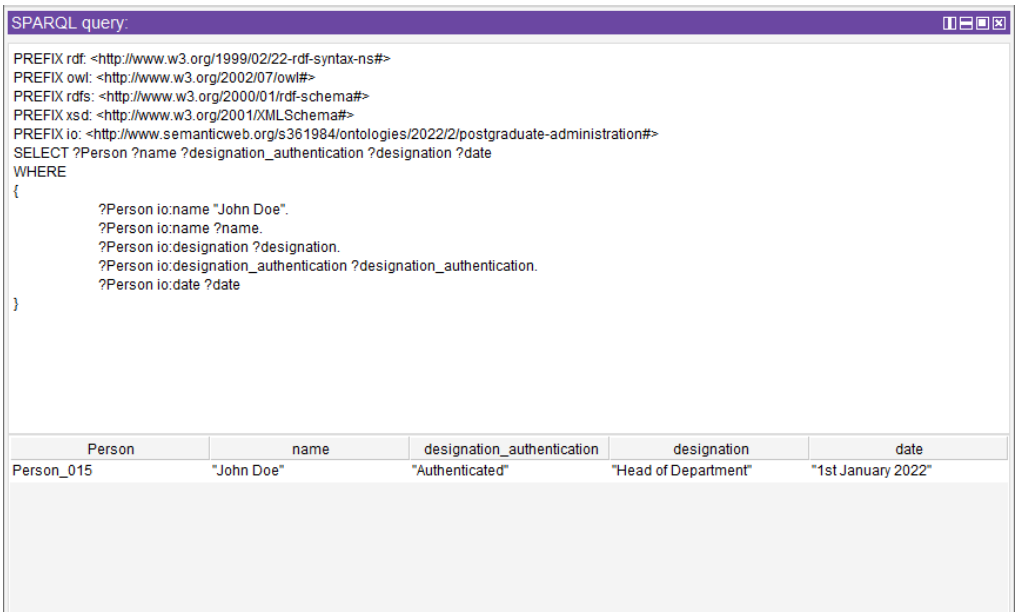

Figure 12: Competency question 5 – Querying personnel's designation authenticity

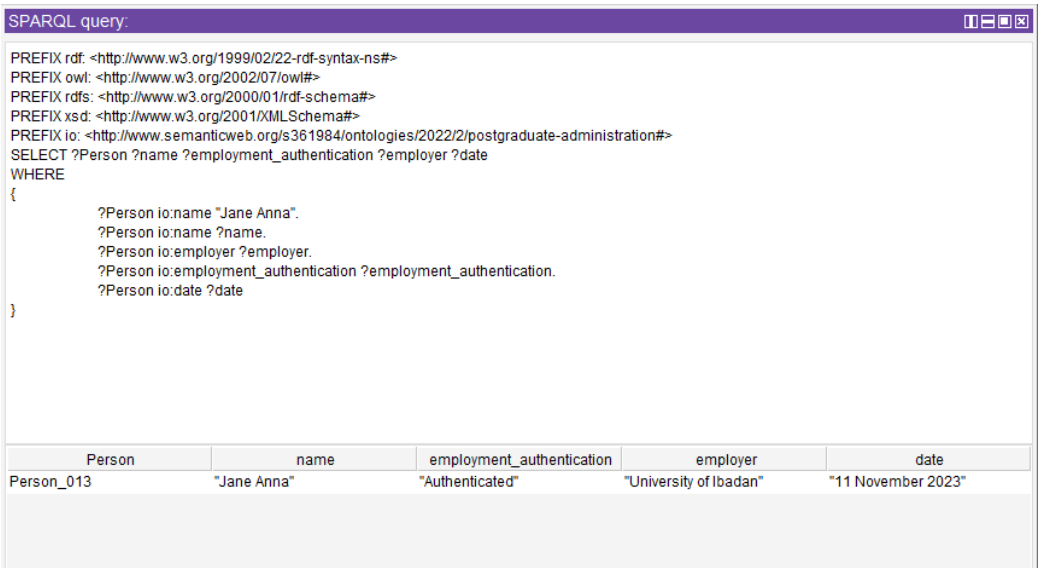

Figure 13: Competency question 6 – Querying personnel's employment authenticity

V. Discussion of Results

The aim of this research was to implement provenance to improve the traceability and authenticity of administration from knowledge representation perspective. Provenance has been applied to previous research domains such as Supply chain [17], Health [13], Climate Study [2] and even Biology [11]. These works made use of Ontology development approach like the one used here, however the domain chosen is quite unique

and different. Also, the use of competency questions made the provenance administration ontology more effective. The different provenance points were validated with the competency questions while the ELK reasoner showed correctness of the ontology in terms of compatibility with the original ontology that was extended. The competency answers showed consistency with the outlined knowledge. Finally, the provenance equation introduced in this research is very valuable because it helps to conceptualise a way to understand how concepts can change from one state to another.

VI. CONCLUSION

In conclusion, the administrative provenance ontology has been built and tested using an existing ontology, description logics, ontology editing software, competency questions and SPARQL queries. In future, an area of interest would be to build different administrative applications around the ontology and benchmark their performances against each other.

REFERENCES

[1] *Philip C. Jackson, Jr. On achieving human-level knowledge representation by developing a natural language of thought. 2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Elenth Annual Meeting of the BICA Society. Procedia Computer Science 190 388–407 (2021)*

[2] *J. Bedia, D. San-Martin, M. Iturbide, S. Herrera, R. Manzanas, J.M. Gutierrez. The METACLIP semantic provenance framework for climate products. Environmental Modelling and Software 119 (2019) 445 –457. (2019)*

[3] *Formal Ontology: Foundation of Domain Knowledge Sharin and Reusing. Journal of Computer Science and Technology. Vol 17. No 5. 535-548. (2002).*

[4] *Anila Sahar Butt(B) and Peter Fitch. ProvAnalyser: A Framework for Scientific Workflows Provenance. 8th International Conference, MODELSWARD 2020 Valletta, Malta, February 25–27(2020).*

[5] *Iman Naja, Milan Markovic, Peter Edwards, and Caitlin Cottrill. A Semantic Framework to Support AI System Accountability and Audit. 18th International Conference, ESWC 2021.Virtual Event, June 6–10, 2021 Proceedings https://doi.org/10.1007/978-3-030-77385-4_10. (2021)*

[6] *Shirly Stephen, Cogan Shimizu, Mark Schildhauer, Rui Zhu, Krzysztof Janowicz, and Pascal Hitzler. A Pattern for Representing Scientific Taxonomies. WOP22: 13th Workshop on Ontology Design and Patterns, October 23-24, 2022, Hangzhou, China. (2022)*

[7] *Yushan Liu, Xiaokui Shu, Yixin Sun, Jiyong Jang, and Prateek Mittal. RAPID: Real-Time Alert Investigation with Context-aware Prioritization for Efficient Threat Discovery. ACSAC '22, December 5–9, 2022, Austin, TX, USA. (2022)*

[8] *Sumaira Manzoor, Yuri Goncalves Rocha, Sung-Hyeon Joo, Sang-Hyeon Bae, Eun-Jin Kim, Kyeong-Jin Joo, and Tae-Yong Kuc. Ontology-Based Knowledge Representation in Robotic Systems: A Survey Oriented toward Applications. Appl. Sci. 2021, 11, 4324. https://doi.org/10.3390/ (2021)*

[9] *Satya S. Sahoo, Matthew D. Turner, Lei Wang, Jose Luis Ambite, Abhishek Appaji, Arcot Rajasekar, Howard M. Lander, Yue Wang, and Jessica A. Turner. NeuroBridge ontology: computable provenance metadata to give the long tail of neuroimaging data a FAIR chance for secondary use. Front. Neuroinform. 17: 1216443.doi: 10.3389/fninf.2023.1216443. © 2023 Sahoo, Turner, Wang, Ambite, (2023)*

[10] *Chuiko, Gennady, Krainyk, Yaroslav, Dvornik, Olga, Darnapuk, Yevhen. IoT in Provenance Management of Medical Data Studies in Computational Intelligence, 941, pp. 347 - 362, Cited 0 times. DOI: 10.1007/978-3-030-64619-6_15. (2021)*

[11] *Yi Chen, Fons.J. Verbeek and Katherine.J. Wolstencroft. FAIR Functional Enrichment: Assessing and Modelling Provenance in Omics Results. The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences. (2023)*

[12] *Cogan Shimizu, Pascal Hitzler and Charles F. Vardeman. A Pattern for Modeling Computational Observations. WOP2022: 13th Workshop on Ontology Design and Patterns, October 23-24, 2022, Hangzhou, China (2022)*

[13] *Dilek Yilmazer Demırel; Özgü Can. ProvVacT: A Provenance Based mHealth Application for Tracking Vaccine History. 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC). (2021)*

[14] *RESEARCH Open Access*

[15] *Max Schröder, Susanne Staehlke, Paul Groth, J. Barbara Nebe, Sascha Spors and Frank Krüger. Structure-based knowledge acquisition from electronic lab notebooks for research data provenance documentation. Journal of Biomedical Semantics (2022).*

[16] *Raúl Ortega1, Enrique Wulff2 & Miguel A. Fortuna. Ontology for the Avida digital evolution platform. https://doi.org/10.1038/s41597-023-02514-3 (2023)*

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

[18] *Monika Solanki, Christopher Brewster. OntoPedigree: Modelling Pedigrees for traceability in supply chains. Aston Business School, Aston University, Birmingham United Kingdom. (2009).*

[19] *Yasmmin Cortes Martins, Artur Ziviani, Maiana de Oliveira Cerqueira e Costa, Maria Cla´udia Reis Cavalcanti, Marisa Fabiana Nicola, and Ana Tereza Ribeiro de Vasconcelos. PPIntegrator: semantic integrative system for protein–protein interaction and application for host–pathogen datasets. Bioinformatics Advances, 2023, vbad067.https://doi.org/10.1093/bioadv/vbad06. (2023)*

[20] *lu, Ruqian & Jin, Zhi. From knowledge-based software engineering to knowware based software engineering. Science in China Series F: Information Sciences. 51. 638-660. 10.1007/s11432-008-0060-y. (2008).*

[21] *Kim, Henry & Laskowski, Marek. Towards an Ontology-Driven Blockchain Design for Supply Chain Provenance. (2016).*

[22] *Oyedeji, Oluseyi Ayodeji & Ayorinde, Ibiyinka. (2023). Development of Administrative Processes Ontology: A Case Study of Postgraduate Education Domain General Terms. International Journal of Computer Applications. 185. 23-30. 10.5120/ijca2023923119.*

