A WebProtégé Plugin for Attesting to the Provenance of Ontologies on the Ethereum Blockchain

Simon Curty[1][0000−0002−2868−9001], Hans-Georg Fill[1][0000−0001−5076−5341], Rafael S. Gonçalves[2][0000−0003−1255−0125], Mark A. Musen[3][0000−0003−3325−793X]

[1] University of Fribourg, Digitalization and Information Systems Research Group
{simon.curty,hans-georg.fill}@unifr.ch
[2] Harvard Medical School, Center for Computational Biomedicine
rafael_goncalves@hms.harvard.edu
[3] Stanford University, Stanford Center for Biomedical Informatics
musen@stanford.edu

Abstract. Ontologies are shared, formal conceptualizations of a domain that are consumed by human and machine agents alike. Trust in ontologies is a central issue for their application. For example, machine learning algorithms for medical diagnosis may rely on the correctness of ontologies and could potentially deliver false results. For enhancing trust, we developed a WebProtégé plugin for the decentralized attestation and verification to the integrity and validity of ontologies using the Ethereum blockchain. Blockchains are an immutable, tamper-resistant and decentralized storage where all transactions are digitally signed. Thus, they permit tracing the provenance of concepts and identify responsible actors. For a first experimental evaluation, we evaluated the transaction costs for attesting to the provenance of ontologies.

Keywords: Ontology · Attestation · Blockchain · Ethereum · WebProtégé

1 Introduction

With the recent integration of linked data, knowledge representations and machine learning in the semantic web, it has become essential both for human and machine agents to know about the provenance of data and derived information [7,5]. Thereby, ontologies play a central role as a formal knowledge resource. Through blockchains as *append-only*, *immutable*, *decentralized* and *distributed* data stores, trust is achieved through *full transparency* of the recorded, *digitally-signed transactions* that are verified through *peer-to-peer consensus protocols*.

In this work we therefore present a plugin for the WebProtégé collaborative ontology editor which enables the attestation to the provenance of ontologies using the Ethereum blockchain. In contrast to the full storage of ontologies on blockchains, which is infeasible due to the limited storage space and comparatively high transaction costs, attestations permit *decentralized*, *verifiable*, and

transparent proofs of the existence and integrity of information without storing the information itself [6]. In addition to its adequacy for blockchain-based applications, attestation further prevents the disclosure of the underlying information, while still allowing for conducting proofs of existence via so-called zero-knowledge-proofs if required [4].

2 Related Work

Traditionally, digital signatures of ontologies have been used for so-called *policy-based trust*. As a foundation for deriving signatures for RDF graphs, Carroll proposed a canonicalization of RDF graphs without changing their semantics [3]. An approach for computing a digest of RDF graphs for content identifiers without the need for canonicalization was discussed by Sayers and Karp [11]. Based on this, Kasten et al. described the signing of individual sub-graphs [10]. However, digital signatures typically rely on a centralized public-key infrastructure whereas blockchains offer a decentralized, distributed, peer-to-peer architecture [14]. Multiple benefits have been previously identified for applying blockchains in the semantic web [2], e.g., for using RDF as the data storage format on blockchains and thus providing a decentralized, immutable, tamper-proof data storage for RDF graphs [12]. Another approach has been proposed in the form of *knowledge blockchains* for the transparent monitoring of ontology evolution and proving the existence of concepts without disclosing them using so-called zero-knowledge proofs [4]. Tuán et al. presented a hybrid approach for storing RDF triplets for use in edge networks, where triplets are stored in a distributed off-chain RDF store but access is controlled by smart contracts [13].

3 Extension of WebProtégé for Enabling Ontology Attestations

The cloud-based, collaborative *WebProtégé* editor is a well-established platform for ontology authoring, which can be extended through a plugin system in the form of portlets [9]. Therefore we chose it as foundation for realizing our implementation for attesting to the provenance of ontologies. The implementation is composed of three major components. A plugin (i) for WebProtégé offers the UI for attesting to or verifying a loaded ontology. For that purpose, a digest of the ontology is calculated by reverting to the OWL API [8]. Further, a transaction containing the attestation information (IRI, version IRI, digest, name of the signer) to the Ethereum blockchain is initiated. The connection and interaction with the chain network is provided by Metamask[4] (ii), a browser wallet extension. Transactions are submitted to the network from a user account, i.e. users are prompted to login with their Ethereum account and authorize the transaction. Thereby, the transaction is digitally signed and sent to a smart contract

[4] Metamask - `https://metamask.io/`

(iii) on the chain. The contract stores the received data, thus irrevocably persisting the attestation information. Finally, a user may query the smart contract to verify if an ontology has been attested to or was changed. The result of the verification presents the user with information on the time and date of the attestation as well as on who attested to the origin of the ontology. In this way it can for example be inferred in a decentralized fashion and without someone having to host the ontology centrally that an expert has validated and attested to the provenance of an ontology, thus increasing the trust that can be placed in that ontology.

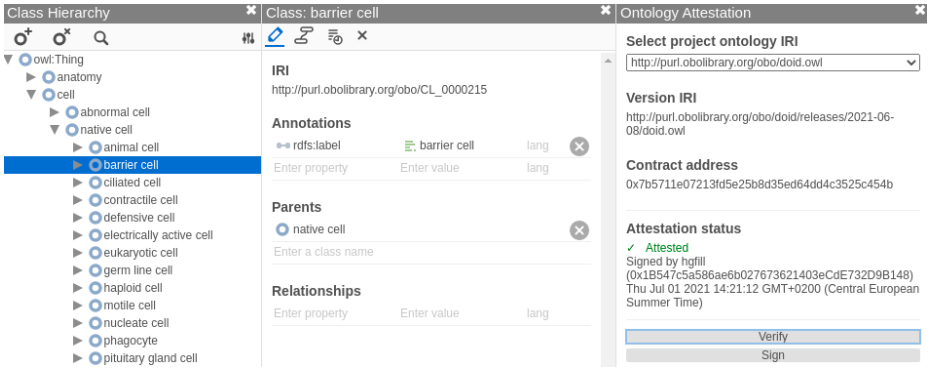

Fig. 1. The attestation portlet integrated into WebProtégé (on the right). A demonstration can be found at `https://www.youtube.com/watch?v=1iuoYz_WTeo`

4 Experimental Evaluation

In Ethereum, size and computational complexity of transactions are measured in an energy unit, called *Gas*. This is a fee payed by the originator in the cryptocurrency Ether and depends on the complexity of commands to be executed [1]. The Ether price of a unit of gas is influenced by the transaction volume. Figure 2 shows historical transaction fees in USD for an attestation in comparison to a baseline, a contract storing a 256bit integer value. E.g., on 1. June 2021, an ontology attestation would have cost USD 8.62 (vs. baseline of USD 3.25). Both gas and ether price are highly volatile. As such, the transaction cost may change significantly in a short period of time.

However, Ethereum is in the process of adopting the proof-of-stake consensus mechanism[5], enabling higher transaction throughput and better energy efficiency. Alternatively, our approach may be adapted to other blockchain-based systems, e.g., Avalanche[6], an Ethereum compatible proof-of-stake blockchain,

[5] See `https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/`
[6] Avalanche - `https://www.avax.network/`

or the Bloxberg infrastructure[7] for decentralized services for the scientific community. We provide the experimental dataset and prototype implementation as open source[8]. The attestation approach, in form of a portlet plugin, has been integrated in a fork of the original WebProtégé distribution. Submitting authorized transactions to the chain network is not handled by the plugin itself, but instead delegated to the required browser wallet extension (Metamask). Thus, users are required to install this extension prior to using the plugin. In return, users retain full control of chain interactions.

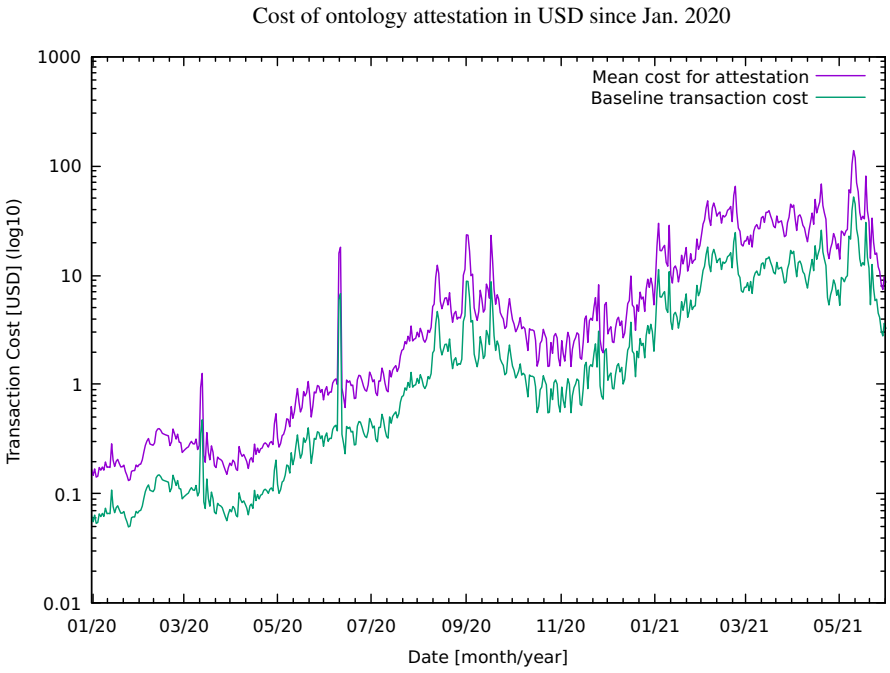

Fig. 2. Transaction costs for attestations based on the price for ETH. The attested ontology only has a minor influence on the incurred costs as the digest length is fixed but the IRI and the signer's name are not.

5 Conclusion

In this paper we described a WebProtégé plugin for attesting to the provenance of ontologies using the Ethereum blockchain. The feasibility of the approach has been evaluated through a prototypical implementation and a cost evaluation.

[7] Bloxberg infrastructure - `https://bloxberg.org/`
[8] Repository - `https://github.com/curtys/webprotege-attestation-base`

Future work will include the investigation of alternative hashing procedures for ontologies for enabling zero-knowledge proofs on a more fine granular level and the extension of the smart contract implementation towards supporting attestations by multiple users.

Acknowledgments

The research on this paper has been partially financed by the Swiss National Science Fund grant number 196889.

References

1. Antonopoulos, A.M., Wood, G.: Mastering ethereum: building smart contracts and dapps. O'reilly Media (2018)
2. Cano-Benito, J., Cimmino, A., García-Castro, R.: Towards blockchain and semantic web. In: Business Information Systems Workshops. pp. 220–231. Springer (2019)
3. Carroll, J.J.: Signing RDF Graphs. In: ISWC. pp. 369–384. Springer (2003)
4. Fill, H.G.: Applying the Concept of Knowledge Blockchains to Ontologies. In: AAAI 2019 Spring Symposium. CEUR-WS.org (2019)
5. Fill, H., Härer, F.: Supporting trust in hybrid intelligence systems using blockchains. In: AAAI 2020 Spring Symposium. CEUR-WS.org (2020)
6. Härer, F., Fill, H.: Decentralized Attestation of Conceptual Models Using the Ethereum Blockchain. IEEE CBI Conference **01**, 104–113 (2019)
7. van Harmelen, F., ten Teije, A.: A boxology of design patterns for hybrid learning and reasoning systems. J. Web Eng. **18**(1-3), 97–124 (2019)
8. Horridge, M.: OWL API main repository. https://github.com/owlcs/owlapi (2020)
9. Horridge, M., Gonçalves, R.S., Nyulas, C.I., Tudorache, T., Musen, M.A.: WebProtégé: A Cloud-Based Ontology Editor. In: World Wide Web Conference. pp. 686–689. ACM (2019)
10. Kasten, A., Scherp, A., Schauß, P.: A Framework for Iterative Signing of Graph Data on the Web. In: ESCW Conference. pp. 146–160. Springer (2014)
11. Sayers, C., Karp, A.: Computing the digest of an RDF graph. Tech. rep., HP Laboratories Palo Alto (2004), `https://www.hpl.hp.com/techreports/2003/HPL-2003-235R1.pdf` (last access 2021-04-09)
12. Sopek, M., et al.: GraphChain: A Distributed Database with Explicit Semantics and Chained RDF Graphs. In: The Web Conference 2018. pp. 1171–1178. ACM (2018)
13. Tuán, A., Hingu, D., Hauswirth, M., Le-Phuoc, D.: Incorporating Blockchain into RDF Store at the Lightweight Edge Devices. In: Int. Conf. on Semantic Systems. pp. 369–375. Springer (2019)
14. Yakubov, A., Shbair, W.M., Wallbom, A., Sanda, D., State, R.: A blockchain-based PKI management framework. In: IEEE/IFIP Network Operations and Management. pp. 1–6. IEEE (2018)