Intelligent Agents: The Vision Revisited

Sabrina Kirrane[1] and Stefan Decker[2]

[1] Vienna University of Economics and Business, Vienna, Austria
[2] RWTH Aachen University, Germany, Germany

Abstract. As early as the mid sixties, motivated by the ever growing body of scientific knowledge, scholars identified the need for data to be organised in a manner that is more intuitive for humans to digest. Additionally, they envisioned a future where intelligent systems would be able to make sense of the data and alleviate humans from performing complex analytical tasks. Although Semantic Web technologies have demonstrated great potential in this regard, the vision has yet to be realised. In this position paper, we highlight some of the challenges experienced by Linked Data publishers who want to make their data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (FAIR), and argue for additional research on policy languages, transparency, trust, and service discovery, with a view to making the intelligent agent vision a reality.

1 Introduction

The idea to use graphs to represent knowledge that can be automatically actioned upon by machines has been around since the early 60's. Both, Engelbart [9] (in 1962) and Lickleder [18, 22] (in 1965) imagined a future where machines would be able to automatically process and reason over data represented in knowledge graphs. Almost forty years later the seminal Semantic Web paper by Berners-Lee et al. [3] described their vision of a Semantic Web, whereby the existing web infrastructure could be used to represent data in a manner that could be automatically integrated and interpreted by software agents.

Roughly five years after the seminal paper both Shadbolt et al. [24] and Feigenbaum et al. [10] reflected on the state of the art at the time and concluded that although intelligent agents were still far from being realised the technology was steadily gaining traction especially as a means of data integration. More recently, Glimm and Stuckenschmidt [13] and Bernstein et al. [4] confirm, that approximately 17 years on, the vision has yet to be realised. The authors observe that although the primary focus was initially on knowledge representation, reasoning and querying, in recent years there has been a broadening beyond pure semantic web topics to include knowledge extraction, discovery, search and retrieval. According to Bernstein et al. [4] there are still a number of issues concerning heterogeneity both in terms of representation and semantics, and diversity in terms of web data quality. Additionally the authors identify new challenges that arise with increasing data volume and publishing velocity.

The goal of this position paper is to revisit the original vision and to highlight open research questions that we deem are important if we hope to one day have intelligent software agents. In particular, (i) we examine current Linked Data publishing practices through the lens of the FAIR[3] (**F**indable, **A**ccessible, **I**nteroperable and **R**eusable) data principles; and (ii) we highlight the need for additional research on policy languages, transparency, trust, and service discovery in order to bring us a step closer to the vision of intelligent software agents.

The remainder of the paper is structured as follows: *Section* 2 introduces the FAIR data principles and discusses some of the limitations of current Linked Data sharing practices. *Section* 3 presents the challenges and opportunities that still need to be overcome for the Semantic Web vision to be realised. Finally *Section* 4 concludes the paper and identifies several open research questions.

2 Making Linked Data FAIR

A recent article by Beek et al. [2] claims that the existing Semantic Web is neither traversable nor machine-processable, and consequently argue that the Semantic Web needs centralisation. In this position paper, we argue for treating the root cause of the problem rather than the symptoms. An emerging best practice in terms of scientific knowledge dissemination is the adoption of FAIR data principles [28], whereby researchers strive to ensure that their research objects (papers, datasets, code etc...) are **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. Although the FAIR principles refer specifically to scholarly data we believe that said principles hold great promise in terms of giving guidance to Linked Data publishers and could improve the quality of Linked Data in general. In order to improve readability in this paper we use (meta)data to denote to data and metadata.

2.1 FAIR Data

The core objective underpinning the FAIR data principles is to provide guidance to data publishers in terms of making their data reusable by both humans and machines. The four foundational principles can be summarised as follows:

- To be deemed **F**indable, (meta)data should be uniquely identifiable via persistent identifiers, these identifiers should be used to associate descriptive metadata with the data, and both data and metadata should be indexed in a manner that is easy to search.
- In order to make data **A**ccessible it should be possible to retrieve the data via common protocol(s), that are open, free, universally implementable and can support usage constraints where desirable.
- Making (meta)data **I**nteroperable is primarily concerned with the representation of (meta)data in a manner that facilitates integration e.g. using common/standard ontologies and vocabularies.
- Finally, (meta)data is **R**eusable if it is richly described in terms of relevant attributes, contains relevant provenance information and is compatible with domain specific standards.

[3] FAIR data practices, https://www.force11.org/node/6062

2.2 FAIR Linked Data

The FAIR data principles are technology agnostic, however nonetheless there is a strong connection between said principles and Semantic Web technologies and Linked Data principles. Both **R**eusability and **I**nteroperablity are at the core of the Resource Description Framework (RDF) data model. By using RDF to describe resources, it is possible to describe complex relations between resources in a machine readable format. Ontologies provide for a shared understanding of things and how they are related, that can easily be reused and extended. Data is linked to other data using HyperText Transfer Protocol (HTTP) Uniform Resource Identifiers (URIs) that can be used to identify things (papers, datasets, code etc...)[4]. Although the RDF data model and Linked Data principles are good starting points in terms of making data FAIR, there are still a number of open research challenges.

In terms of **F**indablility, according to FAIR data and metadata should be identifiable via `persistent identifiers`. Although there is a push by the community to use persistent identifiers, for instance for resources submitted to the International Semantic Web Conference (ISWC) resources track[5], they are not widely used in practice. Another key aspect of **F**indablility is the `indexing` of (meta)data in a manner that is easy to search. Although there have been a number of proposals (cf. [2, 11]), **given that indexing is done in a centralised manner existing proposals suffer from data freshness issues**.

From an **A**ccessible perspective when it comes to `usage constraints` there is a large body of work on access control specification and enforcement strategies for RDF [20] and licensing [14, 15, 16, 26] proposals for data exposed as Linked Data (cf. Section 3 for additional details), however existing **usage control strategies (where used) are still very primitive**.

In this position paper, we argue that the Semantic Web community should **use the FAIR data principles as best practice when it comes to publishing data** and researchers should **use the FAIR principles to develop a strategy for dealing with the root cause of the Findablility and Accessible issues the Semantic Web community need to deal with**.

3 Towards Intelligent Agents

Unfortunately, FAIR does not give guidance on how machines should make use of this vast knowledge base. In this context software agents need to be able to collaborate and potentially negotiate with other agents, bringing with it the need to specify constraints (e.g. in the form of policies), a certain degree of transparency with respect to services offered, the ability to trust other agents and a means to interact and find other agents.

[4] Later the W3C introduced Internationalised Resource Identifiers (IRIs), which provides support for the richer Unicode character set.

[5] http://iswc2018.semanticweb.org/call-for-resources-track-papers/

3.1 Using Policies to Specify Constraints

Berners-Lee et al. [3] originally envisioned a Peer-to-Peer (P2P) network, where each peer is an agent that acts as a data publisher and consumer, and also a service provider and user. One of the key components of such a system is the policy language that is capable of capturing the constraints under which the agents operate. During the early days of the Semantic Web the development of general policy languages that leverage semantic technologies (such as KAoS [7], Rei [19] and Protune [6]), was an active area of research. General policy languages cater for a diverse range of functional requirements (e.g., access control, query answering, service discovery, negotiation, to name but a few). Considering that the policy language needs to be interpreted by machines, formal semantics is important as it allows for the verification of correctness. However, research into general semantic policy languages seems to have reduced considerably in recent years and **the suitability of existing general policy languages towards the intelligent agents vision** is an still open research question.

In terms of specific policy languages access control is a topic that has received a lot of attention over the years. Kirrane et al. [20] provide a detailed survey of the various access control models, standards and policy languages, and the different access control enforcement strategies for RDF. Although there have been several different proposals over the years, there is still no standard access control strategy for Linked Data. Considering the array of access control specification and enforcement mechanisms proposed to date, a necessary first step towards ensuring that intelligent agents have the ability to decide with whom they share information is to develop **a framework that can be used to evaluate existing access control offerings** in terms of expressivity, correctness and completeness.

When it comes to intersection of Semantic Web and licensing, research topics range from using Natural Language Processing to extract license rights and obligations [8] to licenses compatibility validation and composition [14, 15, 16, 26]. More recently, the Open Digital Rights Language (ODRL)[6], which became a W3C recommendation in February 2018, provides a promising first step towards the general adoption of machine understandable licenses, however open research challenges include **the attaching of licenses to data** and **license aware data querying and processing**.

Another promising research direction that remains underdeveloped is the use of policies to specify **norms and values that would enable agents to understand the constraints of the environment** in which they operate. Also, there are also several open research questions in terms of the suitability of the existing languages to **deal with the volume, velocity, variety and veracity of data** we are faced with today, the ability to **balance expressivity and computational complexity**, and ensuring that the intelligent agent ecosystem can deal with the **policy interoperabiliy needs of collaborating agents**.

---

[6] https://www.w3.org/TR/odrl-model/

3.2 Transparency and Trust

In an intelligent agent ecosystem local provenance chains could be used by agents to provide explanations for decisions made, while global provenance chains could be used for transparency with respect to collaborating agents or the distributed system as a whole. These provenance chains could also be used to record and retrieve historical data and to build trust between agents.

To date there has been a number of proposals for representing provenance events (cf. [12, 17]). To date the focus has been on recording where the data came from or capturing the source of the data or changes to data over time. In this regard there have been a number of standardisation initiatives, such as *PROV*[7] and *OWL-Time*[8] ontologies, that can be used to represent *provenance* and *temporal* information respectively. In the context of intelligent agents **there is a need to record provenance with respect to both data and processing in a manner than can be easily digestible**.

From a provenance chains perspective there are two distinct avenues that could be leveraged, one built on top of existing web protocols [23, 27] and another based on blockchain technologies [29]. Weitzner et al. [27] present their vision of a policy-aware architecture for the Web, which includes three basic components: policy-aware audit logging, a policy language framework, and accountability reasoning tools. Specifically, they discuss how transparency and accountability can be achieved via distributed accountability appliances that communicate using existing web protocols. Seneviratne and Kagal [23] build on this idea by proposing a distributed accountability platform known as Accountable Hyper Text Transfer Protocol (HTTPA) that allows data producers to express usage restrictions and data consumers to express usage intentions. Unfortunately the authors only touch upon the required features and **the proposed accountability platform has yet to be assessed from both a functional or a non-functional requirements perspective**. An alternative distributed architecture based on blockchain technology that can be used to provide transparency with respect to personal data processing is proposed by Bonatti et al. [5], however the authors simply describe the opportunities and challenges, and the concrete implementation is left to future work. Zyskind et al. [29] discuss how the blockchain data model and Application Programming Interfaces can be extended to keep track of both data and access transactions. One of the primary drawbacks of the work is the fact that the authors focus on how to repurpose the blockchain as an access-control moderator as opposed to exploring the suitability of the proposed architecture for data transparency and governance. Another related avenue of blockchain research proposes a semantic index for distributed ledgers, which exposes data as Linked Data [25]. Although, Blockchain platforms such as Ethereum[9] and Hyperledger Fabric[10] have the capability to support policy aware service provision, via smart contracts and chaincode, **the suitability of**

[7] PROV,https://www.w3.org/TR/prov-overview/
[8] OWL-Time,https://www.w3.org/TR/owl-time/
[9] https://www.ethereum.org/
[10] https://www.hyperledger.org/projects/fabric

such platforms in terms of interoperability, performance and scalability remains an open research question.

From a trust perspective, Artz and Gil [1] conducted a comprehensive survey of trust mechanisms in computer science in general and the Semantic Web in particular. The authors highlight that traditional approaches focused primarily on authentication via assertions by third parties, however in later years the topic evolved to include historical interaction data, the transfer of trust from trusted entities, and decentralised trust mechanisms (e.g. voting mechanisms or other consensus decision making mechanisms). Although there is a large body of computer science literature relating to trust **the effectiveness of existing trust mechanisms in the context of intelligent agents** has yet to be determined.

Here again research is needed in order to ascertain the **challenges around data volume, velocity, variety and veracity**. In addition, there are a variety of societal challenges that also need to be considered, such as **algorithmic biases, fake news, filter bubbles**, to name but a few.

3.3 Services and Discovery

When it comes to intelligent agents the services offered by each agent need to be designed in such a manner that multiple agents can collaborate in order to complete tasks and solve problems. Each agent needs to maintain a list of services that it is capable of executing based on the (meta)data in its knowledge graph (including descriptive attributes, policies and provenance data). Ideally, the list of services should grow organically with the data and as the agent uncovers new insights based on incremental analysis of its knowledge graph.

Unlike traditional web services, semantic web services use formal ontology-based annotations to describe the service in a manner that can be automatically interpreted by machines. In the early years of the Semantic Web there were several standardisation initiatives, namely the Web Ontology Language for Web Services (OWL-S)[11], the Web Service Modeling Language (WSML)[12], the W3C standard Semantic Annotations for WSDL and XML Schema (SAWSDL)[13]. A survey conducted by Klusch et al. [21] provides a summary of existing work and describes the various semantic web service search architectures (i.e. centralised and decentralised directory based, and decentralised directoryless). The authors conclude that **research into decentalised semantic service search is lagging far behind its centralised counterpart**.

When it comes to semantic web services the big question is how do we support **adaptive discovery and composition of semantic services**. Other open research challenges are concerned with enabling **interoperability between policy aware agents**, and **dealing with agents joining and leaving the P2P network at will**.

[11] https://www.w3.org/Submission/OWL-S/
[12] https://www.w3.org/Submission/WSML/
[13] https://www.w3.org/TR/sawsdl/

4 Conclusion and Future Work

In this paper, we revisit the original vision of the Semantic Web where intelligent agents are able to perform complex computational tasks on behalf of humans [3]. Inspired by recent surveys [4, 10, 13, 24] that analyse the evolution of Semantic Web technologies over almost two decades, we strive to shed light on important research topics in relation to policies, transparency, trust, and service discovery, that although highly relevant are currently under represented at popular international publishing venues for Semantic Web researchers. With this paper we hope to rejuvenate interest in these topics with a view to bringing us closer to making the intelligent agent vision a reality.

References

[1] D. Artz and Y. Gil. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2), 2007.

[2] W. Beek, L. Rietveld, S. Schlobach, and F. van Harmelen. Lod laundromat: Why the semantic web needs centralization (even if we don't like it). *IEEE Internet Computing*, 20(2), 2016.

[3] T. Berners-Lee, J. Hendler, O. Lassila, et al. The semantic web. *Scientific american*, 284(5), 2001.

[4] A. Bernstein, J. Hendler, and N. Noy. A new look at the semantic web. *Communications of the ACM*, 59(9), 2016.

[5] P. Bonatti, S. Kirrane, A. Polleres, and R. Wenning. Transparent personal data processing: The road ahead. In *International Conference on Computer Safety, Reliability, and Security*, 2017.

[6] P. A. Bonatti and D. Olmedilla. Rule-based policy representation and reasoning for the semantic web. In *Proceedings of the Third International Summer School Conference on Reasoning Web*. Springer-Verlag, 2007.

[7] J. M. Bradshaw. *Software agents*. MIT press, 1997.

[8] E. Cabrio, A. Palmero Aprosio, and S. Villata. These Are Your Rights. In *The Semantic Web: Trends and Challenges*, Cham, 2014. Springer International Publishing. ISBN 978-3-319-07443-6.

[9] D. C. Engelbart. Augmenting human intellect: A conceptual framework. summary report afosr-3223 under contract af 49 (638)-1024, sri project 3578 for air force office of scientific research. *Stanford Research Institute. Retrieved March*, 1, 1962.

[10] L. Feigenbaum, I. Herman, T. Hongsermeier, E. Neumann, and S. Stephens. The semantic web in action. *Scientific American*, 297(6), 2007.

[11] J. D. Fernández, W. Beek, M. A. Martínez-Prieto, and M. Arias. Lod-a-lot: A queryable dump of the lod cloud. 2017.

[12] G. Fu, E. Bolton, N. Q. Rosinach, L. I. Furlong, V. Nguyen, A. Sheth, O. Bodenreider, and M. Dumontier. Exposing provenance metadata using different rdf models. *arXiv preprint arXiv:1509.02822*, 2015.

[13] B. Glimm and H. Stuckenschmidt. 15 years of semantic web: An incomplete survey. *KI-Künstliche Intelligenz*, 30(2), 2016.

[14] G. Governatori, A. Rotolo, S. Villata, and F. Gandon. One license to compose them all. In *International Semantic Web Conference*, 2013.

[15] G. Governatori, H.-P. Lam, A. Rotolo, S. Villata, G. A. Atemezing, and F. L. Gandon. Live: a tool for checking licenses compatibility between vocabularies and data. In *International Semantic Web Conference*, 2014.

[16] G. Guido, L. Ho-Pun, R. Antonino, V. Serena, and G. Fabien. Heuristics for Licenses Composition. *Frontiers in Artificial Intelligence and Applications*, 2013.

[17] O. Hartig. Provenance information in the web of data. *LDOW*, 538, 2009.

[18] J. R. Hauben. Vannevar bush and jrc licklider: Libraries of the future 1945–1965. *The Amateur Computerist*, page 36, 2005.

[19] L. Kagal and T. Finin. A policy language for a pervasive computing environment. In *Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks*, pages 63–74. IEEE Comput. Soc, 2003.

[20] S. Kirrane, A. Mileo, and S. Decker. Access control and the resource description framework: A survey. *Semantic Web*, 8(2), 2017. URL `http://www.semantic-web-journal.net/system/files/swj1280.pdf`.

[21] M. Klusch, P. Kapahnke, S. Schulte, F. Lecue, and A. Bernstein. Semantic web service search: a brief survey. *KI-Künstliche Intelligenz*, 30(2), 2016.

[22] J. C. R. Licklider. Libraries of the future. 1965.

[23] O. Seneviratne and L. Kagal. Enabling privacy through transparency. In *Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on*, 2014.

[24] N. Shadbolt, T. Berners-Lee, and W. Hall. The semantic web revisited. *IEEE intelligent systems*, 21(3), 2006.

[25] A. Third and J. Domingue. Linked data indexing of distributed ledgers. In *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017.

[26] S. Villata and F. Gandon. Licenses compatibility and composition in the web of data. In *Third International Workshop on Consuming Linked Data (COLD2012)*, 2012.

[27] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. J. Sussman. Information accountability. *Communications of the ACM*, 51(6), 2008.

[28] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3, 2016.

[29] G. Zyskind, O. Nathan, et al. Decentralizing privacy: Using blockchain to protect personal data. In *Security and Privacy Workshops (SPW), 2015 IEEE*, 2015.