From Data Spaces to Agent Spaces

Axel Polleres[1,2,*,†], Diego Rincon-Yanez[1,†], Amin Anjomshoaa[1,3,†], Daniel Dobriy[1] and Emanuel Sallinger[3,†]

[1] *WU – Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria*
[2] *Complexity Science Hub Vienna, Austria*
[3] *Research Studios Austria (RSA), St. Pölten, Austria*
[3] *TU Wien, Karlsplatz 13, 1040 Vienna, Austria*

Abstract

Data Spaces have been incorporated into Europe's Digital Strategy as an alternative, ledger-based approach to sharing data in a decentralised, self-sovereign manner. As technological developments focus on effective, decentralised data sharing, the integration with (tool-enabled) LLMs and emerging Agentic AI remains unclear. In this position paper, we outline how the core ideas of Data Spaces should be thought forward, towards Agent Spaces. We conclude that such an integration will eventually need to rely on hybrid AI approaches that combine the base idea of self-sovereign data, with (1) self-sovereign Agentic LLMs (incl. tools/services), building on (2) Linked Knowledge Graphs, and adding (3) symbolic guardrails that govern the interaction of agents. Such "bilateral" (agentic and neuro-symbolic) AI approaches integrate symbolic rules (policies and constraints) and planning capabilities with subsymbolic LLMs and learning capabilities.

Keywords

Data Spaces, AI Agents, Semantic Technologies, Large Language Models

1. Introduction

Data spaces are federated data ecosystems built on shared policies, governance rules, and contracts, connecting stakeholders willing to share data in a decentralised yet controllable manner. The main driver behind the idea of data spaces lies in overcoming legal and technical barriers to data sharing while preserving *data sovereignty*. In particular, the European Data Strategy[1] promotes data spaces as a foundation for interoperable and decentralised data sharing across organisational boundaries. As such, Data Spaces complement earlier initiatives towards fueling the data economy, such as Open Data [1]: both Data Spaces and Open Data share for instance the necessity for decentralised *data catalogues*, in order to foster mutual discoverability of data assets and stakeholders; Data Spaces specifically address the requirement that not all data is shareable without constraints amongst these stakeholders, but rather that there is a need for infrastructures enabling them to dynamically define safe data-sharing "spaces". The design of a data space requires addressing both technical and governance aspects. The *technical dimension* encompasses domain-independent services, including metadata management, data discovery, and secure data exchange mechanisms. The *governance dimension* covers operational agreements, access control policies, legal compliance, and shared usage rules among participants.

In the present paper, we aim at placing (i) the current state of Data Spaces initiatives, (ii) their (historical) connection to the "Semantic Web", "Knowledge Graphs" and "Web Services", and (iii) the recent interest in "Agentic AI", side by side, in order to initiate a broader discussion on joint routes ahead, towards what we call *Agent Spaces*.

Fourth International Workshop on Semantics in Dataspaces (SDS 2026), May 10, 2026, Dubrovnik, Croatia

[*] Corresponding author.
[‡] These authors contributed equally.
axel.polleres@wu.ac.at (A. Polleres); diego.rincon@wu.ac.at (D. Rincon-Yanez); amin.anjomshoaa@wu.ac.at (A. Anjomshoaa); daniel.dobriy@wu.ac.at (D. Dobriy); emanuel.sallinger@tuwien.ac.at (E. Sallinger)
0000-0001-5670-1146 (A. Polleres); 0000-0002-8982-1678 (D. Rincon-Yanez); 0000-0001-6277-742X (A. Anjomshoaa); 0000-0001-5242-302X (D. Dobriy); 0000-0001-7441-129X (E. Sallinger)

[1] https://digital-strategy.ec.europa.eu/en/policies/strategy-data

1.1. Data Spaces Initiatives

Several – mainly European – initiatives have been established to operationalise Data Spaces principles in practice. For example, GAIA-X[2] aims to create a federated and secure data infrastructure that shall ensure interoperability, transparency and data sovereignty across European organisations. GAIA-X provides architectural frameworks and technical specifications to enable trusted data sharing across domains: it shall enable the management of policy rules, and provide specifications for decentralised *data catalogues*, *identity management* and further *enabling infrastructures*[3] to advertise and exchange data amongst different stakeholders and enable control/monitoring of data usage (via transfer or interface access), governed by *contracts* (e.g., regulating usage costs, quality guarantees).

The *International Data Spaces Association (IDSA)*[4] translates these principles into concrete technical and governance mechanisms through its IDS Reference Architecture Model (IDS-RAM). The IDS framework defines roles, certification schemes, trust mechanisms, and secure data exchange components (e.g., connectors) that shall enable data providers to retain control over how their data is used.

Summarizing, such initiatives provide structured reference architectures and blueprints that guide the implementation of interoperable, sovereign, and trustworthy data spaces across sectors: data is advertised via metadata and catalogues, enriched with policy and governance information, and made discoverable through search and discovery services, while concrete access shall remain under the publishers' control, in a policy-compliant manner; such *policies* can be specified by the publisher, agreed bilaterally, or overall as part of the governance for the entire space in *contracts* that shall regulate trusted exchange of data within a space.

1.2. The role of Knowledge Graphs and Linked Data within Data Spaces

The core ideas of Data Spaces naturally overlap with and benefit from technologies around (Linked) Knowledge Graphs, that stem from the original "Semantic Web" [2] vision: the Resource Description Format (RDF) specification [3] arguably was designed as a *metadata* standard for describing decentrally managed Web resources, naturally providing decentralised governance through URIs and resp. domain ownership by "namespaces". Building upon this idea, and also using such URIs for "metadata fields", enabled dereferencable vocabularies, discoverable again via URIs, the semantics of which can be axiomatised with RDF's accompanying *RDF Schema* and *OWL* standards. As such, RDF and Linked Data may be viewed as already providing the idea of a shared, decentralised *metadata catalogue* required for Data Spaces.

In parallel, the obvious more general applicability of RDF not only as a *meta*data format, but also as a universal (graph) *data* format, led to the evolution of *Linked Data* [4]: publishing not only metadata but eventually the data itself using RDF, while fully relying on the scale-free self-organisation of the Web, can be viewed as creating a "Giant Global Graph"[5] by dereferencable URI links [5].

Yet, one might argue that this "Giant Global Graph" has not been achieved at Web scale, and we observe two obvious main reasons for this:

Obstacle 1: sensitivity & sovereignty firstly, the very same concerns that prevent universal adoption of Open Data also apply in a Linked Data context: not all data is shareable or combinable with other data assets in an unconstrained manner.

Obstacle 2: scaling semantic agreement secondly, schema heterogeneity does not simply go away by using a syntactically uniform graph data format. That is, reuse and sharing of both instance and schema identifiers needs processes, tools, and services that leverage agreement and mediation.

As a consequence of Obstacle 1, we have seen the evolution and uptake of RDF and Knowledge Graphs (KGs) rather in confined contexts, be it (centrally organised) open collaborative knowledge

[2]https://gaia-x.eu/
[3]potentially, including other base services such ledger-based, decentralised, monitoring/logging capabilities, etc.
[4]https://internationaldataspaces.org/
[5]https://web.archive.org/web/20160713021037/http://dig.csail.mit.edu/breadcrumbs/node/215

graphs [6] (with Wikidata [7] being the prototypical example), or confined (likewise, centrally controlled) enterprise KGs [8].

As an alternative to such centrally controlled KGs, arguably, the core idea of Data Spaces to support the formation of controllable sub-spaces for reasonably sized/scoped groups of stakeholders could be appealing: rather than striving for decentralised agreement on a global scale, agreement between Linked Knowledge Graphs within such "spaces" seems more likely achievable. We herein call such Data Spaces, that rely on linked KGs and semantic agreement, "(Knowledge) Graph Spaces".

In summary, KG(-enabled Data) Spaces, provide, as main advantage (i) a more fluid distinction between meta-data and data, by putting them at the same level, (ii) URIs as a mechanism to maintain links with ownership (by namespaces). Yet, we view KG Spaces as just an intermediate step: Obstacle 1 remains not entirely addressed, since mechanisms for access control in Knowledge Graphs and RDF (e.g., by encryption [9] or query-based access control [10]) remain underdeveloped and not standardised in Graph Processing frameworks (e.g., Triple Stores).

Additionally, especially as for the Obstacle 2, it is still unclear how to automate/scale the formation of semantic agreement, unless provided centrally. What we call here Obstacle 2 directly relates to the challenges *2.1. (Federating Vocabulary Hubs)*, *2.5 (Data Discovery)*, and *2.7. (Schema Alignment and Semantic Data Transformation)* outlined by the W3C Community Group on Dataspaces [11].

Rather than expanding on *all* further challenges from [11] that are currently being actively worked on, we will next discuss extensions of Data Spaces that most closely relate to *2.6. (Pipelining Workflows Across Participants)*: addressing challenge 2.6 requires the consideration of tools and services (Section 1.3) as well as Agentic LLMs (Section 1.4) as first-class citizens within "Data Spaces".

1.3. The role of Tools, Services, Agents within Data Spaces

Dynamic mediation services within Data Spaces (sometimes called intermediary services), just as other *services* like policy engines and contract management services, identity management are already explicitly named components in Data Spaces architectures. Yet, one should not restrict services in Data Spaces to specific components only.

Rather, the consideration of *tools and services* as first-class citizens in Data Spaces seems natural in general. This view is also backed up by another historic development branch of the Web as a whole and semantic technologies in particular: Web Services [12] have been an essential part of the Web's architecture and indeed in many cases the distinction between a data asset and a service (such as a data-providing API) is blurry.

Indeed, the Semantic Web community has since long argued that Web-accessible tools should be considered and made accessible for *automatic composition and invocation* under the line of research related to (Semantic) Web Services [13, 14] advocating catalogues and discovery mechanisms not being restricted to data assets, but rather include tools and services as well. Additionally, automated techniques for composition, orchestration, and choreography of such services have been explored [15]. We suggest to re-consider the original Semantic Web vision [2], based on enabling the seamless collaboration between Services and Agents, rather than mere data sharing only, also within Data Spaces.

Thus, it seems justified to argue towards the evolution of the concepts of Data Spaces, towards "*Service Spaces*" in general. Indeed, the same concepts and considerations regarding Data Assets[6] translate naturally to services: firstly, as in data spaces, the discovery and composability of service assets requires catalogues and meta-data descriptions; secondly, sovereign, controllable aggregation of services in a decentralised manner, will rather be possible in "safe", governed spaces than across the whole Web, in a completely open manner (Obstacle 1 still applies). Lastly, scaling agreement on interfaces remains a barrier in "Service Spaces" (also Obstacle 2 still applies). As such, "Service Spaces" may also be viewed as only an intermediate step.

[6] or, likewise specifically linked, graph data assets as a special case, i.e. what we called "Graph Spaces" above.

1.4. The role of (Agentic) LLMs within Data Spaces

Recent, rapid advances around LLMs have arguably made a sudden leap forward towards implementing the vision of automated tool and services use and composition to achieve complex goals [15]: the models' ability to call and compose tools suggests to also consider (Agentic) LLMs themselves as first-class citizens within the vision of Data Spaces, let us call this latter extension "Agentic Spaces".

In several gradual advancements, Large Language Models (LLMs), based on the idea to leverage a transformer architecture for language generation, have recently advanced far beyond: from the ability to answer questions in natural language, to the generation of executable code, up to learning how to use tools [16], as well as to refine and decompose complex tasks in a (reactive) planning cycle [17]. New protocols and ecosystems have emerged that leverage these capabilities: Anthropic's MCP (model context protocol) [18] or Google's A2A (Agent-to-Agent protocol) [19] allow (i) to register and provide descriptions of executable tools/services and (ii) to enable (read and write) access to data sources for such "Agentic" LLMs. Additionally, "Personal Agent" frameworks such as OpenClaw[7] have further extended this idea towards LLM based agents that keep an own persistent memory and promise to act on behalf of a human, to automatically compose services and data in fulfilment of user-defined goals.

What we herein will call "*Agentic Spaces*" is leveraging such LLM-based agents as actors. Simply put, one could be tempted to argue that allowing agents to freely "roam" within Data Spaces, accessing data catalogues along with a trusted finite set of tools could potentially help address Obstacle 2: Indeed, as we argue in [15], agentic Models can already solve and address a lot of service mediation and data integration problems, based on natural language descriptions and learning from the provision of examples, rather than relying on fully formalised and matched ontologies.

Yet, when considering Obstacle 1 (sensitivity and sovereignty) it is important to also consider the underlying LLMs themselves as part of the trusted space. I.e., also models should be trusted and (self-)sovereign, chosen and in full control of the stakeholders that agree to share a "space".

In summary, for now simply let us define an *Agentic Space* as *a "Data Space" that allows "Agentic Models" to be used within/on behalf of its stakeholders*, for accessing data, services and tools within the Data Space. This may involve tasks such as discovery, service composition, etc. for achieving more complex goals. We note that such a definition has two important implications on Obstacle 1:

Obstacle 1a an "Agentic Space" makes the LLM (and its providers) a **stakeholder** in the data space; we can view this as an extension of Obstacle 1, as it makes the definition of the set of trusted stakeholders in a space more complex; also, mechanisms to monitor stakeholders' data (or other assets') *usage*, probably would need to track *usage by agency*, i.e., which agent used data on which stakeholder's behalf, for which original purpose, etc.

Obstacle 1b governance can only be guaranteed in as far as this new stakeholder also plays by the contracts and policies of the space.

We summarise and illustrate our discussed extensions of the basic "Data Spaces" vision in Figure 1.

Facing the above introduced obstacles, any expectation to simply leverage Agentic AI inline with the core principles of Data Spaces, namely guaranteeing decentralisation and self-sovereignty, seems to be a "red herring": the ongoing discussions about non-deterministic, black-box behaviours of LLMs and biases within these models only amplify these obstacles:

Obstacle 1 As for (1a) Self-sovereign models aren't around the corner, as long as the training, training data and inference provision of models is controlled by a few players (i.e., the providers of API-based models). At the same time the operation of local models, even trained or fine-tuned on controlled data cannot be fully safeguarded: as for (1b), governance policies and guardrails need to take into account potential non-compliance issues, or even deviations from user-defined goals implied by delegation to agentic models.

[7]https://openclaw.ai/

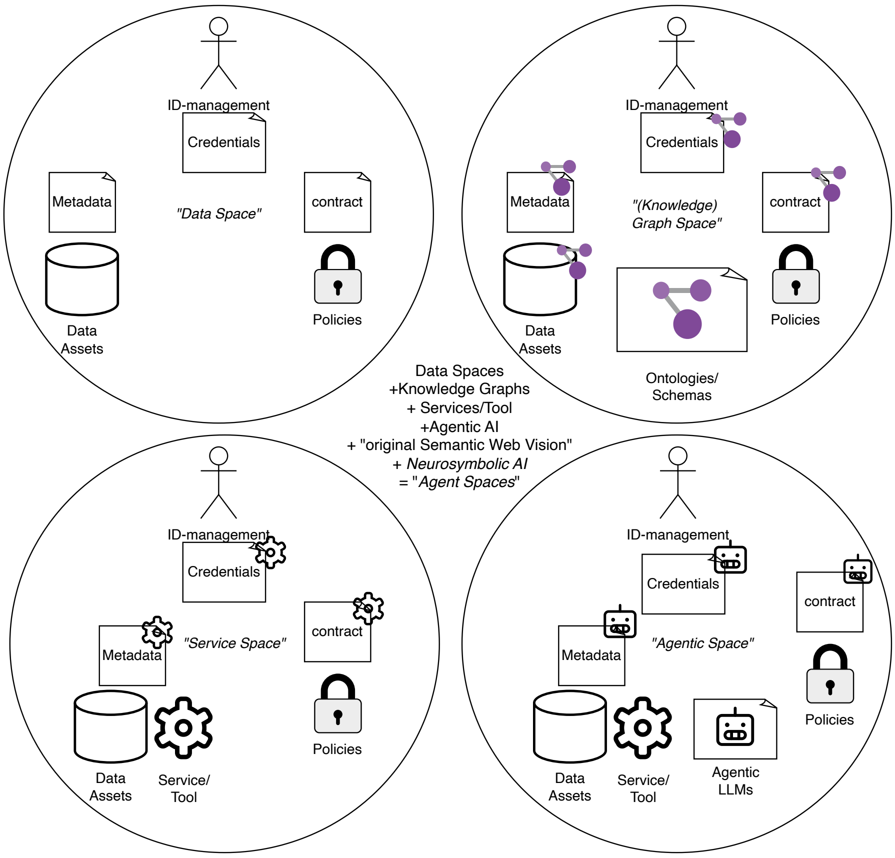

Figure 1: Intermediate Steps towards "Agent Spaces" (Data Spaces, Graph Spaces, Service Spaces, and Agentic Spaces) and their respective. first-class citizens: ID-Management & Credentials, Metadata & Data Assets, Contracts & Policies, Vocabularies (Ontologies/Schemas), Services/Tools, and Agentic LLMs

Obstacle 2 While agentic models can aid in and accelerate overcoming Semantic heterogeneities, there is still a lack of guarantees and explainability of LLM-based mediation.

Hence, in order to distinguish our vision of "real" **Agent Spaces** from what we called "Agentic Spaces", let us define *Agent Spaces* boldly as *"Making the original Semantic Web vision real in a controlled, self-sovereign, decentralised space of Agents"* (humans, organisations, agentic models, tools/services), governed by policies, and *fully overcoming Obstacles 1+2.*

2. Towards Agent Spaces

While at this point, we cannot be sure whether the full of Agent Spaces can be realised, we will present some starting points in the following that

- Combine Data Spaces, KG Spaces, Service Spaces and Agentic Spaces
- Address aspects of Obstacles 1+2.

At the moment, we see the following concrete starting points, all of which build upon the insight that overcoming boundaries of current Agentic AI will require both further advances on the underlying neural architectures, as well as Neuro-Symbolic [20, 21, 22] approaches:

- Building "Agent Spaces" upon *KG Spaces*: (explainable) data catalogues, discovery, and integration based on SPARQL and Agents
- Building "Agent Spaces" upon *Service Spaces*: (i) Symbolic guardrails and process models governing service and agent interaction, (ii) Machine processable, actionable vocabularies for policies
- Building "Agent Spaces" upon *Agentic Spaces*: Self-Sovereign Agentic Models (SLMs).

2.1. Building "Agent Spaces" upon KG Spaces

The Semantic Web standards and technology stack (RDF, SPARQL, along with standardised vocabularies) in combination with agentic AI enables automation of (meta-)data cataloguing, discovery and also dynamic integration of data sources. For instance, we have recently prototypically shown that agentic integration enables both the automatic crawling of SPARQL endpoints and other sources, thus building semantically annotated data catalogues over SPARQL endpoints on the fly: SPARQL-MCP [23] is an open-source server implementing federated (SPARQL 1.1) query processing, accessible to agents via the MCP protocol. As such, it is a first step in am Agent Spaces architecture geared towards enabling AI Agents to query structured data and metadate, bu providing capabilities for AI agents to generate and execute federated queries. Additionally, SPARQL-MCP enables schema discovery and exploration through automated VoiD [24] schema description retrieval and extraction components. While still in a prototype stage, such automated `schema retrieval` capabilities, in combination with further agentic tools for partially automating schema alignment, will be particularly suitable for deployments in scoped agent Spaces where they can leverage and accelerate semantic agreement.

The global scalability problem of automated agents exploring Web sources such as SPARQL endpoints on the fly [25] may be less of an issue in confined Data Spaces, plus also in such spaces, common policies for meta-data publishing may be easier to establish than on the Open Web.

Beyond automating data access, also both symbolic repairs [26, 27], as well as KG Embeddings and Graph Neural Networks (GNNs) [28] could be used within Data Spaces to predict and reconcile inconsistencies and missing information.

2.2. Building "Agent Spaces" upon Service Spaces

As discussed in a recent invited contribution to the Web Conference's "History of the Web" track [15], the advent of Agentic AI may well revive some of the ideas and concepts around "Semantic Web Services": in order to compose services and tools, we need trusted repositories; likewise, composition based on symbolic planning and scheduling could be paired with the current purely reactive and non-deterministic choice of tools invoked by Agentic LLMs. Indeed, symbolically controlled RAG pipelines that involve explainable interleaving of LLM and service calls may yield more robust (neuro-symbolic) integrations than purely LLM-based agents. Also, apart from their work on services, the Semantic Web and Agents' communities have been working actively over the past few years on symbolic inference techniques that shall enable the automatic verification and enforcement of contracts and policies [29]; such techniques are needed for governing interactions of human and LLM-based agents within Agent Spaces. Also, this work is supported by specific, actively worked on vocabularies such as ODRL [30] and DPV [31] to describe policies in a symbolic, unambiguous manner.

2.3. Building "Agent Spaces" upon Agentic Spaces

Finally, there is a challenge in terms of enabling the "self-sovereign LLMs" (SLMs) necessary to drive what we call Agent Spaces. Such SLMs should be both trainable and operable in a sovereign manner within confined data spaces, or even at the level of single stakeholders/publishers within a data space.

Recent advances around model distillation [32], more efficient fine-tuning [33], and novel architectures such as xLSTMs [34] could make such smaller, more inference-efficient models feasible. LLM distillation to XLSTMs [35] is actively pursued at the moment in the "Bilateral AI" (BILAI) [22] project. Also, there is active work on decentralised training platforms that could potentially "democratise" model training.[8]

In the context of agentic capabilities of such SLMs, we observe that current small, open LLMs still perform poorer than large foundation models in terms of interoperability with structured data in our own experiments with SPARQL-MCP [23] mentioned above in Section 2.1; yet, this is not necessarily due to their limitations in contextual understanding - but due to the lack of such SLMs, fine-tunable to structured data or SPARQL query generation.

[8]to name an example without the intent of advertisement: https://www.tplr.ai/

2.4. A conceptual architecture sketch

Figure 2 illustrates a – deliberately conceptual – sketch on how Linked Knowledge Graphs, Services, and decentrally interconnected agentic SLMs could be combined in an "Agent Space": each stakeholder/organisation runs their own, sovereign LLMs (SLM), based on decentralised knowledge and policies. The interlinked KGs both integrate and virtualise local datasets, and also comprise a virtual, decentralised metadata catalogue: the small circles within the local KGs denote metadata shared across the whole space; consequently, the metadata catalog itself can be generated and hosted as a shared service, for instance by an agent running tools like SPARQL-MCP (cf. Section 2.1).

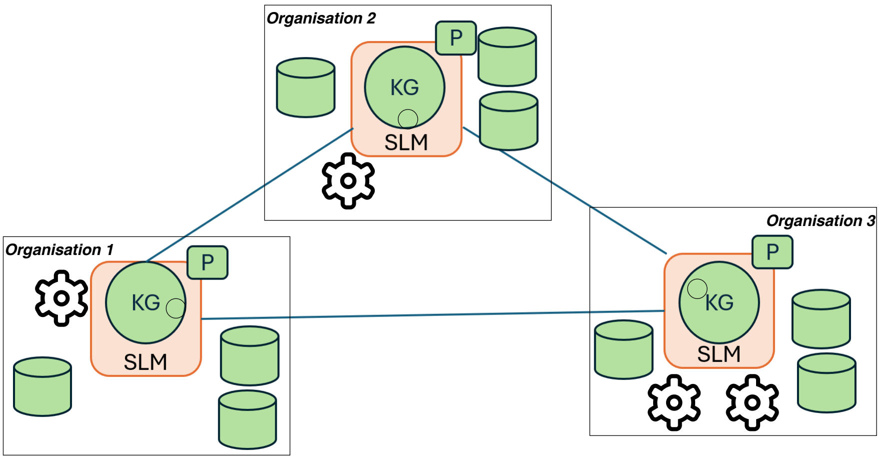

Figure 2: An **Agent Space** with Linked Knowledge Graphs (**KG**s), wrapped by a sovereign, agentic language model (**SLM**), that is governed by symbolic policy guardrails (**P**).

As such, we envision "bilateral"[22] (aka neuro-symbolic) interactions that support such an architecture: firstly, KGs serve as combined data and meta-data (inner circles in Figure 2) providers. Secondly, agentic SLMs (surrounding boxes) can assist to scalably access and reconcile inconsistencies and ontological mismatches in the KGs: starting from RAG, MCP or similar pipelines as the most obvious bilateral integration paradigm, further "bilateral" reconciliation techniques have been outlined in Section 2.1.

Similarly, bilateral interactions are required for a trustworthy policy component (marked by "P" in Figure 2): policies should be controllable and enforceable by deterministic, symbolic rules on the one hand, but must be additionally governed by (user-)explainable, natural language contracts on the other hand: bilateral methods are needed to link these two. Lastly, bilateral combinations of symbolic planning, scheduling, and configuration techniques with "agentic reasoning" patterns (e.g., ReACT [17]) shall determine the course of agents' actions in a trustworthy manner [15].

3. Conclusions

In the present position paper, we postulate that the core principles of sovereignty around Data Spaces should not only focus on data assets, but also consider Knowledge Graphs (*KG Spaces*), Services&Tools (*Service Spaces*, as well as Agentic LLMs (*Agentic Spaces*) as first-class citizens.

We propose to establish the full potential of this combination under the term *Agent Spaces*:[9]

1. leveraging **Bilateral AI** (i.e., neuro-symbolic) capabilities, derived from the combination of self-sovereign Agentic LLMs (SLMs) and Knowledge Graphs (KGs), governed by **symbolic** policy guardrails and constraints
2. focused on a **decentralised** environment akin to Data Spaces.

[9]as a route forward towards the often (over-)stressed "full potential" of the Semantic Web [2]

We have discussed substantial challenges, that should be understood as a "call to action" before we arrive at Agent Spaces: techniques to guarantee scalable, sovereign, controlled KG data access, in terms of *standardised* and agreed means for harnessing query-based access control and encryption within KGs (Obstacle 1); provisioning of LLMs that can savely operate within "spaces" in a sovereign manner (Obstacle 1a+1b); automating and scaling semantic agreement within such "spaces" (Obstacle 2). We discussed (i) agentic data access on top of SPARQL, (ii) more affordable training (distillation/fine-tuning) and provisioning of models, but also more general (iii) bilateral approaches for policy-based governance of self-sovereign Agentic LLMs in such spaces as a potential, non-exhaustive list of starting points. We welcome discussion and more concrete architecture proposals, based on these starting points.

Acknowledgements & AI Declaration. The work on this paper has been funded by the Austrian Science Fund (FWF) 10.55776/COE12. Emanuel Sallinger's work on this paper was also funded by by the Vienna Science and Technology Fund (WWTF) [Grant ID: 10.47379/VRG18013, 10.47379/ICT25032, 10.47379/NXT22018, 10.47379/ICT2201, 10.47379/DCDH001]. The authors have not used Generative AI tools in the preparation of this paper.

References

[1] A. Zuiderwijk, M. Janssen, Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View, Springer New York, New York, NY, 2014, pp. 115–135.

[2] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American (2001).

[3] O. Lassila, R. R. Swick, Resource description framework (rdf) model and syntax specification, 1999. URL: https://www.w3.org/TR/1999/REC-rdf-syntax-19990222, w3C Recommendation.

[4] T. Berners-Lee, Linked Data, 2006-2009. URL: https://www.w3.org/DesignIssues/LinkedData.

[5] A. Haller, J. D. Fernández, M. R. Kamdar, A. Polleres, What are links in linked open data? a characterization and evaluation of links between knowledge graphs on the web, ACM Journal of Data and Information Quality (JDIQ) 2 (2020) 1––34. doi:10.1145/3369875.

[6] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, R. Navigli, A.-C. N. Ngomo, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Computing Surveys (CSUR) 54 (2021) 1–37. Extended pre-print available at https://arxiv.org/abs/2003.02320.

[7] D. Vrandecic, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (2014) 78–85.

[8] J. Sequeda, O. Lassila, Designing and building enterprise knowledge graphs, Springer Nature, 2022.

[9] J. D. Fernández, S. Kirrane, A. Polleres, S. Steyskal, HDT_{crypt}: Compression and Encryption of RDF Datasets, SWJ 11 (2020) 337–359. URL: http://semantic-web-journal.org/content/hdt-crypt-compression-and-encryption-rdf-datasets. doi:10.3233/SW-180335.

[10] S. Kirrane, A. Mileo, A. Polleres, S. Decker, Query based access control for linked data, CoRR abs/2007.00461 (2020). URL: https://arxiv.org/abs/2007.00461. arXiv:2007.00461.

[11] P. Colpaert et al., Dataspace challenges, 2026. URL: https://web.archive.org/web/20260314174255/https://w3c-cg.github.io/dataspaces/, Draft Community Group Report.

[12] G. Alonso, F. Casati, H. Kuno, V. Machiraju, Web Services: Concepts, Architectures and Applications, Springer, 2004.

[13] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGuinness, B. Parsia, T. Payne, M. Sabou, M. Solanki, N. Srinivasan, K. Sycara, Bringing semantics to web services: The owl-s approach, in: Proceedings of the 1st International Workshop on Semantic Web Services and Web Process Composition (SWSWPC), 2004, pp. 26–42.

[14] D. Fensel, H. Lausen, A. Polleres, J. De Bruijn, M. Stollberg, D. Roman, J. Domingue, Enabling semantic web services, Springer Science & Business Media, 2006.

[15] A. Polleres, F. Bauer, D. Dobriy, T. Käfer, L. Kubelka, A. Harth, T. Wehr, On the historic roots of Agentic AI in Semantic Web Services, in: The ACM Web Conference 2026, 2026.

[16] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, E. Hambro, L. Zettlemoyer, N. Cancedda, T. Scialom, Toolformer: Language models can teach themselves to use tools, in: Proceedings of the 37th Annual Conference on Neural Information Processing Systems (NeurIPS), 2023.
[17] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. R. Narasimhan, Y. Cao, React: Synergizing reasoning and acting in language models, in: Proceedings of the 11th International Conference on Learning Representations (ICLR), 2023.
[18] Anthropic, Model context protocol specification, version 2025-06-18, MCP Specification, 2025. URL: https://modelcontextprotocol.io/specification/2025-06-18.
[19] Google, A2A Project Contributors, Agent2agent (a2a) protocol specification, https://a2a-protocol.org/v0.3.0/specification/, 2025. Version 0.3.0.
[20] A. Sheth, K. Roy, M. Gaur, Neurosymbolic artificial intelligence (why, what, and how), IEEE Intelligent Systems 38 (2023) 56–62.
[21] L. De Smet, L. De Raedt, Defining neurosymbolic ai, arXiv preprint arXiv:2507.11127 (2025).
[22] B. A. C. of Excellence, Broad ai: Bilateral ai cluster of excellence – research modules, 2025. URL: https://www.bilateral-ai.net/research/research-modules.
[23] D. Dobriy, F. Bauer, A. Azzam, D. Banerjee, A. Polleres, Agentic SPARQL: Evaluating SPARQL-MCP-powered Intelligent Agents on the Federated KGQA Benchmark, Technical Report, Vienna Unversity of Economics and Business, 2026. doi:10.57938/83c86964-2d48-46f1-b655-5bef78c1a837.
[24] K. Alexander, R. Cyganiak, M. Hausenblas, J. Zhao, Describing linked datasets with the void vocabulary (2011).
[25] R. Taelman, E. Crum, Did ai crawlers kill sparql federation?, in: International Workshop on Knowledge Graphs for Sustainability, 2026. To appear.
[26] N. Ferranti, D. Guimaraes, J. Souza, A. Polleres, Formalizing repairs for wikidata constraint violations: A taxonomy and empirical analysis, in: Proceedings of the 24th International Semantic Web Conference (ISWC 2025), Springer, Nara, Japan, 2025.
[27] R. David, S. Ahmetaj, M. Šimkus, A. Polleres, Repairing SHACL constraint violations using answer set programming, in: Proceedings of the 21st International Semantic Web Conference (ISWC 2022), volume 13489, Springer, Virtual Conference (Hangzhou, China), 2022, pp. 375–391.
[28] M. Vázquez, K. Innerebner, A. Prock, G. Klambauer, E. Lex, J. Schimunek, A. Polleres, Structure is the signal: Graph encodings and GNNs for constraint repair in collaborative KGs, in: 23rd European Semantic Web Conference (ESWC), 2026.
[29] T. Kampik, A. Mansour, O. Boissier, S. Kirrane, J. Padget, T. R. Payne, M. P. Singh, V. Tamma, A. Zimmermann, Governance of autonomous agents on the web: Challenges and opportunities, ACM Trans. Internet Technol. 22 (2022).
[30] A. Cimmino, N. Fornara, V. Rodríguez-Doncel, J. Domingue, International workshop on ODRL and beyond: Practical applications and challenges for policy-base access and usage control (OPAL2025), in: Proceedings of the ESWC 2025 Workshops and Tutorials, volume 3977 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, p. 2. URL: https://ceur-ws.org/Vol-3977/OPAL2025-preface.pdf.
[31] H. J. Pandit, B. Esteves, G. P. Krog, P. Ryan, D. Golpayegani, J. Flake, Data privacy vocabulary (DPV) - version 2.0, in: The Semantic Web - ISWC 2024 - 23rd International Semantic Web Conference, volume 15233 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 171–193.
[32] G. E. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, CoRR (2015).
[33] T. Dettmers, A. Pagnoni, A. Holtzman, L. Zettlemoyer, Qlora: efficient finetuning of quantized llms, in: Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23, 2023, p. 28.
[34] M. Beck, K. Pöppel, P. Lippe, R. Kurle, P. M. Blies, G. Klambauer, S. Böck, S. Hochreiter, xLSTM 7B: A recurrent LLM for fast and efficient inference, in: Forty-second International Conference on Machine Learning, ICML Vancouver, BC, Canada, volume 267 of *PMLR*, 2025, p. 23.
[35] L. Hauzenberger, N. Schmidinger, T. Schmied, A.-R. Hartl, D. Stap, P.-J. Hoedt, M. Beck, S. Böck, G. Klambauer, S. Hochreiter, Effective distillation to hybrid xlstm architectures, CoRR (2026).