

A Categorization of Cross-Domain Semantic Interoperability Challenges for Open (Government) Data

Maria Ioanna Maratsi, Yannis Charalabidis and Charalampos Alexopoulos

University of the Aegean, University Hill, Mytilene, 81100, Greece

Abstract

The benefits of open data have by now been extensively investigated. However, the generated data grows at a rapid pace and, with it, a plethora of concerns including, among others, data quality, interpretability, machine-readability, and interoperability. The present study aims to review recent literature regarding the semantic interoperability challenges for open (and government) data and to examine the status for difficult areas of research which have not yet been sufficiently addressed. The literature review revealed various sector-specific but also general, cross-domain challenges, which were then categorized into four groups according to the source where the issue usually stems from. As far as the status of approaches to the identified challenges is concerned, there appears to be a tendency to avoid fit-to-all solutions and instead follow a more domain-specific strategy to enable semantic interoperability, and allow for cross-domain reuse, wherever this is possible.

Keywords

Semantic interoperability, open data, open government data (OGD), technological challenges

1. Introduction

The European Union Directive 2019/1024 [1] on open data and their re-use, emphasizes the need for all EU Member States to be active and involved in preparations for their infrastructure to support open data as a concept and adopt open and re-use policies for the data generated. The Directive thus pinpoints that the data made available for reuse, as well as the relevant metadata derived from it, needs to be interpretable and machine-readable in order to satisfy the important condition of data interoperability. According to the National Interoperability Framework Observatory (NIFO), the model which describes the most important aspects of interoperability and integrates the concept of interoperability-by-design, consists of six layers: four layers of interoperability (technical, semantic, organizational, and legal interoperability), one component for all four layers which is “integrated public service governance” and one background layer, which is “interoperability governance” [2]. The scope of this research is to focus on the semantic interoperability layer. Interoperability is a multi-dimensional challenge, which, to be addressed to a satisfactory degree, needs improved awareness raising and knowledge relevant to all six layers mentioned previously in the interoperability-by-design paradigm. In this light, the main aim of the presented research is twofold; first, to understand the current situation of existing semantic interoperability challenges and, secondly, to organize the identified areas of semantic challenges using a sectoral approach.

The methodology of this study consisted of the following approach. Initially, the literature was retrieved mainly from the digital libraries of Scopus and IEEE Xplore, limited to the ones with publication date between 2018 and 2022, English language, and using search queries with logical operators and the following keywords: “semantic”, “challenges”, “open data”, “interoperability”, “open government data”. As a next step, the retrieved results were subject to the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA) method [3]. Finally, the information obtained was synthetized in a conceptual categorization of the identified challenges.

2. Findings

Challenges related to semantic interoperability are supported by the conducted structured literature review and presented in Section 2.1.

2.1. Literature Review

The sectors referred to in the retrieved literature were Health, Education, Cultural Heritage, Digital Government, Agriculture, Environment, and Open Statistical Data, the latter being a horizontal, cross-sectoral category. Some indicative results are presented in this Section. For instance, semantic interoperability challenges in the health sector include the heterogeneity of information systems [4], the data fragmentation over multiple silos, the varying medical consent policies, and different legislations on national level [11, 6, 7, 8, 10], but also the risk of faulty information, which information representation models entail, and the inability to express biomedical knowledge in a formal but also straightforward manner [9]. The legislation and policies incompatibility [15], data inconsistencies, semi-structured information, data heterogeneity, and the lack of common vocabularies also concern the digital government sector. Moreover, the need for stable, governed data standards [15], the integration of data sources to represent relationships and allow for cross-domain usage, but also the requirement of collaboration between human and non-human agents are some more aspects of semantic interoperability in this context [16, 17, 18]. Heterogeneous information systems, unstructured and unlinked data are challenges also present in the sector of agriculture [21, 22]. The educational sector, as listed in the examined literature, suffers mainly from poor metadata and the difficulty of linking the data. There is a need for more domain-specific ontologies to formally express knowledge, but also to build on unified education vocabularies, while the varying levels of access to technological means (and thus e-education) is another soft challenge [13, 5, 12, 14]. Similarly, the poor or non-standardized metadata hinders the interoperability of cultural objects, while the lack of accessibility to shared information and formal knowledge representation are strongly present in the cultural heritage sector [19]. Another challenging aspect of semantic interoperability which is frequently encountered in knowledge domains with strong terminology compounds, such as the environment sector, derives from the multidisciplinarity of information, the large variation in technical language, and the different terms of use, conventions, etc. In addition, documentation (guides, manuals etc.) usually exists in various formats and forms (such as plain text), while there is an incessant need for consistent access to methods and practices [20]. Lastly, a cross-sectoral category (as the data might concern all of the aforementioned and more), the open statistical data, suffers from the challenges which disparate sources and portals entail, while another issue of pivotal importance is the difficulty of differentiation between two components which have, in fact, very close semantic proximity but are modeled using different standards [23, 24, 25, 26].

2.1.1. A Categorization of Semantic Interoperability Challenges

All things considered, interoperability challenges can be horizontal and existent in all sectors, while others are more sector specific. Common cases include, among others, the complexity of data integration, the heterogeneous nature of information systems and data, the varying levels of access to technological means by various stakeholders, incomplete, redundant or unstructured data, and the lack of standardized approaches. Focusing specifically on the semantic interoperability challenges, one can see all related issues eventually converging and stemming from common origins. Table 2 presents the identified semantic interoperability challenges, conceptually organized in four categories.

Table 2

A Categorization of Semantic Interoperability Challenges for OGD

Ontological	Metadata-Related	Core Components	Other
Inability to express knowledge in a formal but also straightforward manner.	Need for enrichment of poor metadata.	Information models represent artifacts in which information is assigned, perils of faulty information.	Varying legislation or policy at national level.
Need for definition of domain-specific ontologies.	Lack of semantic relationships to link data.	Unstructured data. Fragmentation.	Lack of new tools and practices and lack of standardization policy.
Need for definition of common, multi-domain, core ontologies.	Non standardized metadata.	Different models for components with close semantic proximity.	Data heterogeneity.
Difficulty to make lightweight ontologies.	Discoverability not enabled.	Lack of common vocabularies.	Domain multidisciplinarity, semantic proximity and difficult differentiation among terms.

3. Conclusions

The purpose of this study was to identify the semantic interoperability challenges for open (and government) data and to examine the status for difficult areas of research which have not yet been sufficiently addressed. The literature review revealed various sector-specific but also cross-domain challenges, conceptually categorized into four groups. The presented work is, however, subject to some limitations. Due to the vastness of the topic and the plethora of existing literature, this study is not exhaustive, and an alternative methodology could be fruitful to extract more insight and complement the literature findings, e.g., expand the knowledge sphere by conducting semi-structured interviews with experts for each domain (e.g., platform developers, industry professionals, government entities, semantic web experts and more) in order to gain perspective of the state-of-the-art in this regard but also potentially identify the stakeholders directly or indirectly affected by these technological difficulties, aiming to emphasize alignment with a more citizen-driven but also data-driven approach.

4. Acknowledgements

The authors acknowledge the financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955569.

5. References

- [1] [1] Official Journal of the European Union, Directive (EU) 2019/1024 of the European Parliament and of the Council, "Open data and the re-use of public sector information", June 2019
- [2] [2] 6 interoperability layers. Joinup. (n.d.). Retrieved November 21, 2022, from <https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/solution/eif-toolbox/6-interoperability-layers>
- [3] [3] Prisma. PRISMA. (n.d.). Retrieved November 21, 2022, from <https://prisma-statement.org/>
- [4] [4] Rahman, H, Hussain, and MI. "A comprehensive survey on semantic interoperability for Internet of Things: State-of-the-art and research challenges". Trans Emerging Tel Tech. 2020; 31:e3902. <https://doi.org/10.1002/ett.3902>
- [5] [5] S. D. Nagowah, H. B. Sta and B. A. Gobin-Rahimbux, "Towards Achieving Semantic Interoperability in an IoT-enabled Smart Campus." 2019 IEEE International Smart Cities Conference (ISC2), 2019, pp. 593-598, doi: 10.1109/ISC246665.2019.9071694.

- [6] [6] European Commission, Directorate-General for the Information Society and Media, Virtanen, M., Ustun, B., Rodrigues, J., et al., “Semantic interoperability for better health and safer healthcare: deployment and research roadmap for Europe”, Stroetmann, V.(editor), Publications Office, 2013, <https://data.europa.eu/doi/10.2759/38514>
- [7] [7] Queralt-Rosinach, N., Kaliyaperumal, R., Bernabé, C.H. et al. “Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic.” *J Biomed Semant* 13, 12 .2022. <https://doi.org/10.1186/s13326-022-00263-7>
- [8] [8] A. N. Gohar, S. A. Abdelmawgoud and M. S. Farhan, “A Patient-Centric Healthcare Framework Reference Architecture for Better Semantic Interoperability Based on Blockchain, Cloud, and IoT.” in *IEEE Access*, vol. 10, pp. 92137-92157, 2022, doi: 10.1109/ACCESS.2022.3202902.
- [9] [9] R. G. Sonkamble, S. P. Phansalkar, V. M. Potdar and A. M. Bongale, “Survey of Interoperability in Electronic Health Records Management and Proposed Blockchain Based Framework: MyBlockEHR.” in *IEEE Access*, vol. 9, pp. 158367-158401, 2021, doi: 10.1109/ACCESS.2021.3129284.
- [10] [10] N. ANGULA, N. DLODLO and P. Q. MTSHALI, “Enabling Semantic Interoperability of Crowdsourced Disease Surveillance Data for Namibia Through a Health-Standards-Based Approach.” 2019 IST-Africa Week Conference (IST-Africa), 2019, pp. 1-9, doi: 10.23919/ISTAFRICA.2019.8764830.
- [11] [11] F. N. S. Palma, “Interoperability Challenges and Critical Success Factors in the Deployment of Cross-border Digital Medical Prescriptions in Finland and Estonia.” 2022 IEEE International Conference on Digital Health (ICDH), 2022, pp. 60-65, doi: 10.1109/ICDH55609.2022.00018.
- [12] [12] Bashir, Faiza, and Nosheen Fatima Warraich. “Systematic literature review of Semantic Web for distance learning.” *Interactive Learning Environments* (2020): 1-17.
- [13] [13] C. K. Pereira, S. W. M. Siqueira, B. P. Nunes and S. Dietze, “Linked Data in Education: A Survey and a Synthesis of Actual Research and Future Challenges.” in *IEEE Transactions on Learning Technologies*, vol. 11, no. 3, pp. 400-412, 1 July-Sept. 2018, doi: 10.1109/TLT.2017.2787659.
- [14] [14] C. -M. Chituc, “Interoperability Standards in the IoT-enabled Future Learning Environments: An analysis of the challenges for seamless communication.” 2020 13th International Conference on Communications (COMM), 2020, pp. 417-422, doi: 10.1109/COMM48946.2020.9141959.
- [15] [15] Buyle, Raf & Vanlithout, Ziggy & Coetzee, Serena & De Paepe, Dieter & Van Compernolle, Mathias & Thijs, Geert & Nuffelen, Bert & Vocht, Laurens & Mechant, Peter & Vidts, Björn & Mannens, Erik. 2018. “Raising interoperability among base registries: The evolution of the Linked Base Registry for addresses in Flanders. *Journal of Web Semantics.*” 55. 10.1016/j.websem.2018.10.003.
- [16] [16] Tshering, Younten, and Chutiporn Anutariya. “Enabling Semantic Interoperability in Bhutan's E-Government: An Ontology-based Framework.” 2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2022.
- [17] [17] Masoumi, Hadi, Bahar Farahani, and Fereidoon Shams Aliee. “Systematic and ontology-based approach to interoperable cross-domain open government data services.” *Transforming Government: People, Process and Policy* (2021).
- [18] [18] Rocha, Bartira, et al. “A Linked Data-based semantic information model for smart cities.” 2019 IX Brazilian Symposium on Computing Systems Engineering (SBESC). IEEE, 2019.
- [19] [19] I. Koch, “Integration of models for linked data in cultural heritage and contributions to the FAIR principles.” 2022 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2022, pp. 1-2.
- [20] [20] P. L. Buttigieg, S. Caltagirone, P. Simpson and J. S. Pearlman, “The Ocean Best Practices System - Supporting a Transparent and Accessible Ocean.” *OCEANS 2019 MTS/IEEE SEATTLE*, 2019, pp. 1-5, doi: 10.23919/OCEANS40490.2019.8962680.
- [21] [21] D. Zeginis et al. 2021. “Statistical Challenges Towards a Semantic Model for Precision Agriculture and Precision Livestock Farming.”
- [22] [22] Devare, Medha & Aubert, Céline & Benites-Alfaro, Omar & Masias, Ivan & Laporte, Marie-Angélique. 2021. “AgroFIMS: A Tool to Enable Digital Collection of Standards-Compliant FAIR Data. *Frontiers in Sustainable Food Systems.*” 5. 726646. 10.3389/fsufs.2021.726646.

- [23] [23] Evangelos Kalampokis, Areti Karamanou, Konstantinos Tarabanis. "Towards Interoperable Open Statistical Data. 18th International Conference on Electronic Government (EGOV).", Sep 2019, San Benedetto del Tronto, Italy. pp.180-191, ff10.1007/978-3-030-27325-5_14ff. fffhal-02445809
- [24] [24] Efthimios Tambouris, Evangelos Kalampokis, and Konstantinos Tarabanis. 2017. "Visualizing Linked Open Statistical Data to Support Public Administration." In Proceedings of dg.o '17, Staten Island, NY, USA, June 07-09, 2017, 6 pages. DOI: [hp://dx.doi.org/10.1145/3085228.3085304](http://dx.doi.org/10.1145/3085228.3085304)
- [25] [25] E. Chaniotaki, E. Kalampokis, E. Tambouris, K. Tarabanis and A. Stasis 2017. "Exploiting Linked Statistical Data in Public Administration: The Case of the Greek Ministry of Administrative Reconstruction." 23rd Americas Conference on Information Systems (AMCIS2017)
- [26] [26] E. Kalampokis, A. Karamanou and K. Tarabanis 2019. "Interoperability Conflicts in Linked Open Statistical Data."