

# Utilization of Semantic Technologies to Achieve Interoperability Between Dental Data and EHRs: A Proof of Concept

Hugo Lebredo<sup>1</sup>, Rubén del Rey Álvarez<sup>2</sup> and José Emilio Labra Gayo<sup>3</sup>

<sup>4</sup>Department of Computer Science, University of Oviedo, Oviedo 33007, Spain

## Abstract

The fragmentation of a patient's clinical history between the systems of the different professionals and health providers from different medical areas, both public and private, which they have attended is a common scenario. Several problems arise because of this, such as the appearance of information silos, data incompatibilities between different institutions, duplication of information, and data dispersion.

This work presents a proof of concept that enables individuals to comprehensively manage their dental clinical information, allowing them to share it with other stakeholders according to their preferences. The main feature of this proof of concept is its focus on interoperability, achieved through the integration of semantic technologies, such as Shape Expressions, and widely adopted standards in the medical community, such as FHIR and SNOMED. Furthermore, it aligns with the SOLID (Social Linked Data) principles, promoting a decentralized, secure, and accessible personal clinical data ecosystem.

## Keywords

RDF, ShEx, Decentralization, Interoperability, Personal Health Records, Standalone

## 1. Introduction

Many contemporary health information systems are designed in such a way that domain concepts intended to be processed by the system are hard-coded directly into their software and database models. This design limitation results in the inability of these systems to utilize information from other ecosystems, despite technological advancements that enable inter-system communication. Consequently, this leads to the formation of information silos and the associated issues, including data fragmentation, outdated and duplicated records. Furthermore, information isolation poses a significant challenge, particularly when attempting to integrate data across different medical specialties. The unique characteristics of odontology have led to its recognition as an independent medical discipline in the United States, owing to distinct differences in training, services provided, and financing when compared to general medicine [1]. The lack of time, space, and incompatibility of medical records—both between dental clinics and with medical institutions from other disciplines, such as general medicine—constitutes significant barriers within this medical field. Notably, 91% of active dentists work in private practice settings, with 46% operating in solo practice [2]. Integrating dental records with those from other medical specialties offers numerous benefits for patients. Such integration provides critical insights for the early diagnosis and monitoring of over 200 systemic diseases. It also supports investigative and forensic studies across institutions and allows dental clinics to improve the accuracy of clinical data obtained through patient interviews (anamnesis), thereby enhancing diagnostic decision-making and overall patient care [3]. To resolve the problems described above, we present a proof of concept based on the decentralized architecture proposed by Lebredo et al. [4] which allows to computational agents independent of any stakeholder in charge of the exchange of information between organizations and patients, being able to recover, validate and transform the data in a transparent way and respecting the right of consent for the use of patient information. This architecture combines the use of the SOLID

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025

✉ lebredohugo@uniovi.es (H. Lebredo); rubdelrey@protonmail.com (R. d. R. Álvarez); labra@uniovi.es (J. E. L. Gayo)

🌐 <https://labra.weso.es/> (J. E. L. Gayo)

🆔 0000-0002-8745-2338 (H. Lebredo); 0009-0008-7677-3209 (R. d. R. Álvarez); 0000-0001-8907-5348 (J. E. L. Gayo)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

paradigm (Social Linked Data) [5], semantic web technologies along with ontologies and medical data standards. Its main feature is the storage of clinical information in a single repository under the control of the patient, independent of any medical institution. To achieve this, the healthcare provider's system must develop a software layer that performs two key tasks. First, it must be capable of transforming the clinical information intended to be sent to the patient's personal repository into an RDF graph, along with a Shape Expression [6] that defines the topology of the data contained in the graph. Second, to access the information stored in the patient's personal repository, this layer must validate the data in the Shape Expression against its associated graph and transform it according to the data model used by the domain. This approach allows the data topology to be defined in a transparent manner. Besides this, the information stored into the graph will follow the FHIR<sup>1</sup> standard along with the SNOMED<sup>2</sup> terminology that allow any computational agent to properly consume the information. The granularity of this architecture allows small health providers to exchange clinical information with their patients without the need for expensive developments.

## 2. Decentralised Architecture

We propose a proof of concept that achieves the interoperability of Electronic Health Records (EHR) from heterogeneous data sources applying the SOLID principles (Social Linked data), which uses principles of RDF semantic technologies and the concept of Personal Online Datastore (POD) in conjunction with the Personal Health Records (PHR) paradigm [7]. With this approach, each individual will have full control over their unified health information stored in their own POD, which provides them with the ability to accurately manage access control to that information by third parties.

In this ecosystem, two different types of domains must coexist. First, the organizational domain, to which the providers and health professionals belong. Its purpose is to incorporate new information derived from their professional activity into their systems using EHR. The second type of domain is the personal domain. Its objective is to allow the owners of the data (patients) to access and manage all their medical data in a single repository, even when this data was generated by different actors.

All the information will be stored in the system using Resource Description Format (RDF). Its simple structure makes it an ideal candidate for representing data, metadata, terminology, and ontologies from heterogeneous data sources. RDF has been proclaimed as "the best available candidate for a Universal Health Exchange language" in the Yosemite Manifesto [8]. Interoperability between systems will be achieved by using clinical standards such as FHIR and SNOMED. Both standards can be expressed in RDF and are well-known in data representation and interchange health environments. They allow to homogenize the representation of clinical concepts such as weight, height or blood pressure and combine them creating more complex entities called archetypes such as a cardiovascular report in a format that facilitates its reuse in different systems [9].

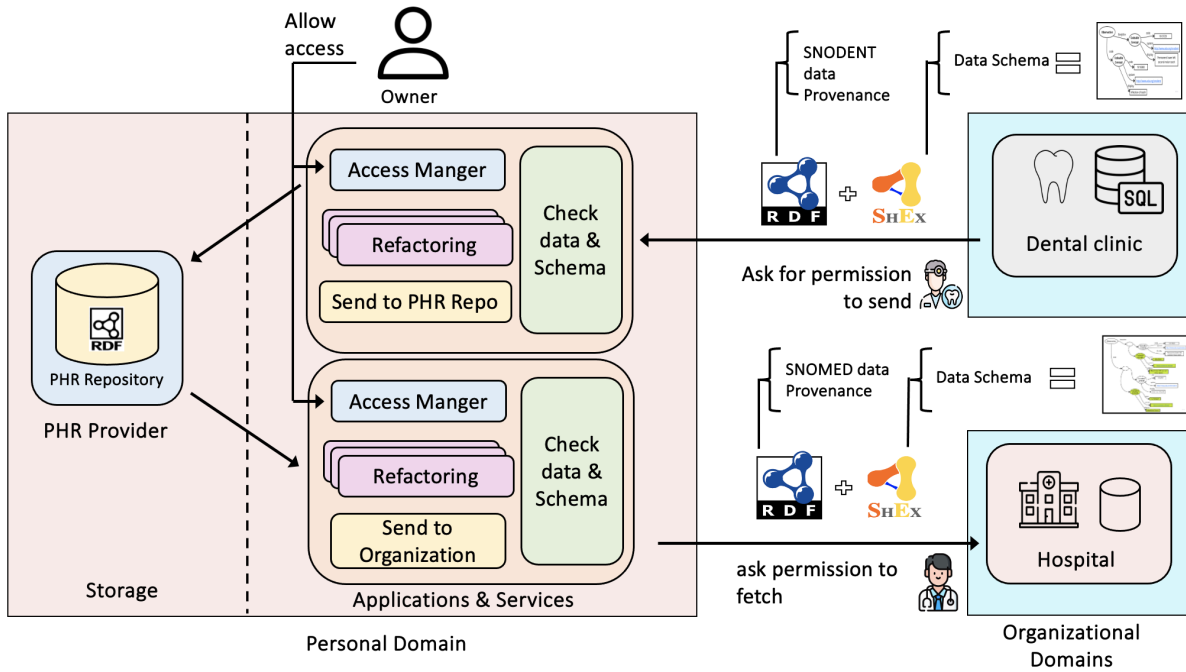
To integrate in a single POD different medical sources with potentially different data schema, we propose to use Shape Expressions (ShEx)[6]. ShEx is a language that allows to describe and validate RDF graphs in a formal way. In a system such as the one that we propose, a Shape Expression can be used in several ways with different purposes:

- *Data integration*: the different medical actors able to generate data that will be stored in a patient's POD may use different internal data schemas. Those schemas could be expressed in ShEx, so our system is able to understand and potentially transform the input data by mapping the provided schema to a different one. Changes performed in the input data due to these mappings could happen at a structural level (reorganizing the information's topology) or at a vocabulary level (changing some entities/properties by some others belonging to standard vocabularies). Also, the schemas allow to automatically validate that the data provided by the medical actors is actually organized as it is declared in the corresponding schema.

---

<sup>1</sup><https://www.hl7.org/fhir/>

<sup>2</sup><https://www.snomed.org/>



**Figure 1:** Data flow and services provided by independent agents.

- *Data accessing:* When a medical actor wants to access a patient's POD, it may not need to query the whole information about a patient, but only the subset of it which is relevant for a certain case. Regardless of the user interface used by the practitioner to express which subset of the information should be retrieved, ShEx could be used to describe the schema of the data that should be retrieved from a patient's POD. A ShEx schema describing the structure and properties of the desired information allows not only to find the data matching such structure, but also to detect what type of data is being queried so the POD's owners could receive precise alerts according to this information to properly decide if they allow access to that specific subset of their information.
- *Data description:* ShEx can be used to expose in an abstract manner the kind of data potentially queryable in each POD. By abstract we mean without including any actual data of any patient, as the ShEx schemas will merely expose structural information instead of actual records. Such schemas could be parsed by automatic agents for data querying and mapping purposes at different levels of our architecture.

Shapes Expressions have already been used to describe and validate FHIR archetypes in medical environments [10, 11].

Figure 1 shows the information exchanged between personal and organizational domains and drafts the communication interfaces between the different agents in this decentralised architecture.

### 3. Proof of Concept

This proof of concept consists of two organizational environments capable of sending and receiving data from a personal POD. One of them simulates a dental clinic, and the other one simulates a hospital, as well as a personal environment composed of a POD server and an application that allows the owner to manage and visualize the content of their POD.

In the dental clinic environment, information related to any intervention stored in its database is capable of being sent to a personal POD. Before this, the system transforms the data into an RDF graph. The triples conform to the FHIR and SNOMED standards to represent clinical information.

Additionally, a Shape Expression is sent, defining the archetype used to describe the data's topology, enabling third-party systems to validate the schema against the associated graph and consume the information contained within the data.

In this proof of concept, within the personal environment, the POD owner can visualize the data provided by professional environments and manage access control through the standard access control lists of PODs.

The second organizational environment can consume information from a patient's personal POD, adapt that information to its own data model using the Shape Expression schema attached to the data, and send clinical information generated by its system.

The two professional domains used in this proof of concept employ different vocabularies. The dental clinic uses the Systematized Nomenclature of Dentistry (SNODENT)<sup>3</sup> vocabulary, maintained and distributed by the American Dental Association (ADA), while the hospital uses the SNOMED<sup>4</sup> vocabulary, which is more prominent in the field of general medicine.

## 4. Conclusions

We propose a proof of concept to integrate medical data from different health providers in a decentralized structure, shifting data ownership from health providers to patients. This prototype, based on a decentralized architecture, has successfully achieved interoperability between data generated by different agents with potentially diverse topologies, allowing it to be combined and queried seamlessly. We propose using RDF graphs to store the data, alongside Shape Expressions to describe and validate it during various processes. SOLID PODs are utilized to implement data ownership and access control, while mappings to well-known medical standards, including FHIR and SNOMED, are employed to ensure structural and vocabulary unification.

The granularity of this architecture, based on decoupling information from the applications that wish to consume it, enables scalability tailored to the needs of different stakeholders. The personal domain plays a significant role, as various applications can interact with the repository's information if the owner grants the necessary permissions.

The proof of concept is available in this GitHub repository [12]

## 5. Future Work

The proof of concept described in this document should be evaluated in future work by developing a more sophisticated prototype that allows for the exchange of more complex clinical information from heterogeneous sources between practitioners and patients, and evaluates its performance. The main challenges of this approach include defining data mappings that are general enough to handle and integrate data from a broad range of use cases, creating a new business layer to manage the hosting of repositories, and ensuring the acceptance of this paradigm shift by both organizations and patients.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

- [1] M. Vujicic, Time for Dental Care to Be Considered Essential in US Health Care Policy, *AMA Journal of Ethics* 24 (2022) E57–63. URL: <https://journalofethics.ama-assn.org/>

---

<sup>3</sup><https://www.ada.org/resources/practice/dental-standards/snodent>

<sup>4</sup><http://snomed.info/sct>

article/time-dental-care-be-considered-essential-us-health-care-policy/2022-01. doi:10.1001/amajethics.2022.57.

- [2] J. L. Fellows, K. A. Atchison, J. Chaffin, E. M. Chávez, N. Tinanoff, Oral Health in America, *The Journal of the American Dental Association* 153 (2022) 601–609. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0002817722002112>. doi:10.1016/j.ada.j.2022.04.002.
- [3] M. H. Torres-Urquidy, V. Powell, S.-M. R. Y. Geist, S. Mishra, M. Chaudhari, M. Allen, Health Information Technology Considerations of Medical and Dental Data Integration, in: A. Acharya, V. Powell, M. H. Torres-Urquidy, R. H. Posteraro, T. P. Thyvalikakath (Eds.), *Integration of Medical and Dental Care and Patient Data*, Springer International Publishing, Cham, 2019, pp. 155–206. doi:10.1007/978-3-319-98298-4\_11, series Title: Health Informatics.
- [4] H. Lebrede, D. Fernández-Álvarez, J. E. Labra-Gayo, A Decentralized Architecture for Semantic Interoperability of Personal Dental Data Based on FHIR, *International Journal on Semantic Web and Information Systems* 19 (2023) 1–16. URL: <https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJSWIS.333633>. doi:10.4018/IJSWIS.333633.
- [5] A. V. Sambra, E. Mansour, S. Hawke, M. Zereba, N. Greco, A. Ghanem, D. Zagidulin, A. Abounaga, T. Berners-Lee, *Solid: A Platform for Decentralized Social Applications Based on Linked Data* (2016).
- [6] E. Prud'hommeaux, J. E. Labra Gayo, H. Solbrig, Shape expressions: an RDF validation and transformation language, in: *Proceedings of the 10th International Conference on Semantic Systems*, ACM, Leipzig Germany, 2014, pp. 32–40. URL: <https://dl.acm.org/doi/10.1145/2660517.2660523>. doi:10.1145/2660517.2660523.
- [7] A. Roehrs, C. A. da Costa, R. d. R. Righi, K. S. F. de Oliveira, Personal Health Records: A Systematic Literature Review, *Journal of Medical Internet Research* 19 (2017) e13. URL: <http://www.jmir.org/2017/1/e13/>. doi:10.2196/jmir.5876.
- [8] Yosemite Manifesto, 2013. URL: <http://www.yosemitemanifesto.org/>.
- [9] L. Marco-Ruiz, D. Moner, J. A. Maldonado, N. Kolstrup, J. G. Bellika, Archetype-based data warehouse environment to enable the reuse of electronic health record data, *International Journal of Medical Informatics* 84 (2015) 702–714. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1386505615300058>. doi:10.1016/j.ijmedinf.2015.05.016.
- [10] D. K. Sharma, E. Prud'hommeaux, D. Booth, C. Nanjo, G. Jiang, Shape Expressions (ShEx) schemas for the FHIR R5 specification, *Journal of Biomedical Informatics* 148 (2023) 104534. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1532046423002551>. doi:10.1016/j.jbi.2023.104534.
- [11] H. R. Solbrig, E. Prud'hommeaux, G. Grieve, L. McKenzie, J. C. Mandel, D. K. Sharma, G. Jiang, Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx), *Journal of Biomedical Informatics* 67 (2017) 90–100. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1532046417300345>. doi:10.1016/j.jbi.2017.02.009.
- [12] R. del Rey, Prototipo tfg, <https://github.com/RubendelRey/Prototipo-TFG>, 2024. Último acceso: 19-12-2024.