International Journal on Semantic Web and Information Systems
Volume 19 • Issue 1

A Decentralized Architecture for Semantic Interoperability of Personal Dental Data Based on FHIR

Hugo Lebredo, University of Oviedo, Spain*
https://orcid.org/0000-0002-8745-2338

Daniel Fernández-Álvarez, University of Oviedo, Spain

Jose Emilio Labra-Gayo, University of Oviedo, Spain
https://orcid.org/0000-0001-8907-5348

ABSTRACT

Several problems arise due to the differences between dentistry and general medicine. The storage of dental data in information silos, the incompatibility of data between different dental clinics or institutions from other medical areas are the most significant ones. The authors propose a decentralized architecture that combines FHIR archetypes, shape expressions, and personal online datastores (PODs) to tackle those issues as follows: FHIR archetypes are used to express the data, shape expressions are used to handle data structure and data access requests, and PODs are used to store information in a decentralized and safe manner that let the owner of the information stored to handle data access. The system allows the patient to store dental information from heterogeneous data sources transparently and respecting the patient's right to autonomy and consent. In this paper, the authors develop this architecture proposal and discuss its relevance and feasibility in the area of dental health.

KEYWORDS

Decentralization, HL7 Fast Healthcare Interoperability Resources (FHIR), Odontology, Personal Health Records, Privacy, RDF, Semantic Interoperability, Shape Expressions (ShEx)

INTRODUCTION

Since the Obama administration declared the development of Electronic Health Records (EHRs) a strategic policy in 2009 and allocated $27 billion to develop such a strategy[1], there has been a rapid digital transformation in healthcare systems on a global scale. Some other government bodies, such as the European Union, have also made progress in adopting similar lines of action[2].

All these strategies are covered by the 2030 Agenda in its Goal 3 "Good health and well-being" and Goal 9 "Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation." They also ratified the suitability of the states that have developed public health policies whose strategic lines are based on the digitization of medical information and encouraged other states to follow this path.

DOI: 10.4018/IJSWIS.333633
*Corresponding Author

Similar suggestions are mentioned in the "Global Strategy on Digital Health 2020-2025" published by The WHO (World Health Organization)[3]. Such a guide recognizes the existence of a broad consensus around the idea that information technologies will be of the utmost importance to improve the health and well-being of 3 billion people through the interaction of three main actors: private sector, civil society, and technical communities in information and communication technologies. The following are strategic objectives:

- Promoting global collaboration and making progress with the transfer of knowledge on Digital Health.
- Advocating digitally enabled, people-centered health systems.
- Advancing the national implementation of Digital Health Strategies.
- Strengthening governance for digital health at global, national, and regional levels.
- Establishing national interoperable digital health ecosystems, as well as strengthening coordinated collaboration, promoting the use of big data and Artificial Intelligence under appropriate ethical principles and a review of regulations.

Each state has its dynamics and its context. Trying to adapt these general strategic lines to a particular reality can be accomplished thanks to the application of public health policies. It is common for governments to articulate the rules and regulations that satisfy the needs of the states themselves and encourage the development of solutions for third parties, imposing the framework of work that health providers and citizens must follow. This is a very complex process that requires a large amount of economic, administrative, technological, and political effort.

For example, the Spanish and French health systems have very notable differences, while still being neighboring countries and having a similar culture. Spain has a very powerful public network of hospitals and medical centers that is fully paid for with citizens' taxes, while France has a smaller public network and has co-payment financing between the state, insurers, and patients as a cornerstone (Chevreul et al., 2015), which gives greater freedom to their citizens to choose the practitioner by whom they want to be treated.

The proliferation of digitization and cloud computing made the development of huge state networks that managed the clinical information of citizens possible. Taiwan's healthcare system can provide a useful frame of reference for possible future developments due to some of its main features. It provides great coverage of patient care ranging from traditional Chinese medicine to odontology, through a cloud information system that allows the integration of medical records used in hospitals and other health care providers. Physicians can access all the information of a patient generated by different institutions, to improve the service provided and avoid duplication in the prescription of medications and performance of treatments (Yeh & Saltman, 2019). For this purpose, a VPN network through which care providers connect their systems with social security information repositories was implemented. Practitioners must use their username and password to access the patient's medical information. To maintain the privacy of particularly sensitive data, patients have the right to set their passwords to protect access to such information. If a physician wanted to see said information, the patient himself would have to enter his password (Lee et al., 2022).

However, these systems have limitations due to the centralized nature of the information. The definition of the scope of the different medical areas that the system will cover, the data model used, and even its availability are decisions that are made unilaterally by policymakers. The problem of information silos, although to a lesser extent, is still present since only a part of the medical information generated in the interventions of the doctor's system is stored in the central repository. In the orthodontic area, the following information is stored: dental treatment, medical treatment interval, medical order amount, and graphic options, which include display according to specific dental treatments or dental annotations.

Clinics, practitioners, and small hospitals use the Cloud information system to a lesser extent than large medical providers, as they cannot afford the expense of developing the integration of their system with the Cloud information system despite having developed a web platform. To solve such an issue, there is a need to invest time in filling the data that small providers do not have.

The particularities of odontology mean it is being treated in the USA as an independent area of medicine, due to the existing differences in terms of training, services provided, and financing concerning general medicine (Vujicic, 2022). In the field of odontology, 91% of active dentists worked in private practice settings and 46% of them were in solo practice (Fellows et al., 2022). The lack of time, space, and incompatibility of medical records are barriers derived from this scenario that professionals face in their day-to-day lives. For these reasons, dental information has many possibilities of being stored in information silos.

In this paper, we propose an alternative approach that allows achieving interoperability between different medical institutions and practitioners. The exchange of dental records with other areas of medicine provides remarkable benefits for patients and professionals. Dental records could provide invaluable information for early diagnosis and monitoring the progress of more than 200 systemic diseases. Additionally, different institutions can use this information for investigative and forensic purposes. Furthermore, dental clinics and other areas will increase the accuracy of clinical information collected by anamnesis; the method by which physicians gather information about a patient's past and present medical condition through questions to make informed clinical decisions. This improves decision-making in diagnostics (Torres-Urquidy et al., 2019).

Electronic Health Records vs. Personal Health Records

Both Electronic Health Records (EHR) and Personal Health Records (PHR) are used to store health-related information. However, there is a conceptual difference between them. The former was designed so that health professionals can perform their work, while the latter is intended for an individual to be able to monitor themselves and share their health data with third parties—normally health professionals, although they can also be other people. To achieve their purposes, important challenges must be faced, such as privacy management and data interoperability.

EHR can be defined as "A repository of information regarding the health of a subject of care in computer processable form, stored and transmitted securely, and accessible by multiple authorized users. It has a commonly agreed logical information model which is independent of EHR systems. Its primary purpose is the support of continuing, efficient and quality integrated health care and it contains information which is retrospective, concurrent and prospective" (ISO 20514, 2005). The EHR includes all the data necessary for a practitioner or healthcare provider to realize their professional activity. That is the medical data of a patient, such as lists of allergies, medication, clinical history, and the administrative information derived from those operations necessary for the management of the provider.

On the other hand, the definition of PHR is "an Internet-based set of tools that allows people to access and coordinate their lifelong health information and make appropriate parts of it available to those who need it" (Connecting for Health. The personal health working group final report, 2003). If we change "people" to "subject of care," which is a synonym for "patient," the spectrum of data covered is broadened to anything that is linked to well-being throughout a person's life. Two other fundamental differences must be mentioned. The first is that PHRs are not simple repositories, but online toolkits. The second one is that the responsibility for data management rests with individuals rather than professionals.

In a PHR-oriented application, different stakeholders interact with different degrees of responsibility. Individuals and physicians are the most relevant actors interacting with the system, but we must remember the importance of others such as non-health professionals, managers, policy makers, and IT professionals, which are necessary for the development and maintenance of the application.

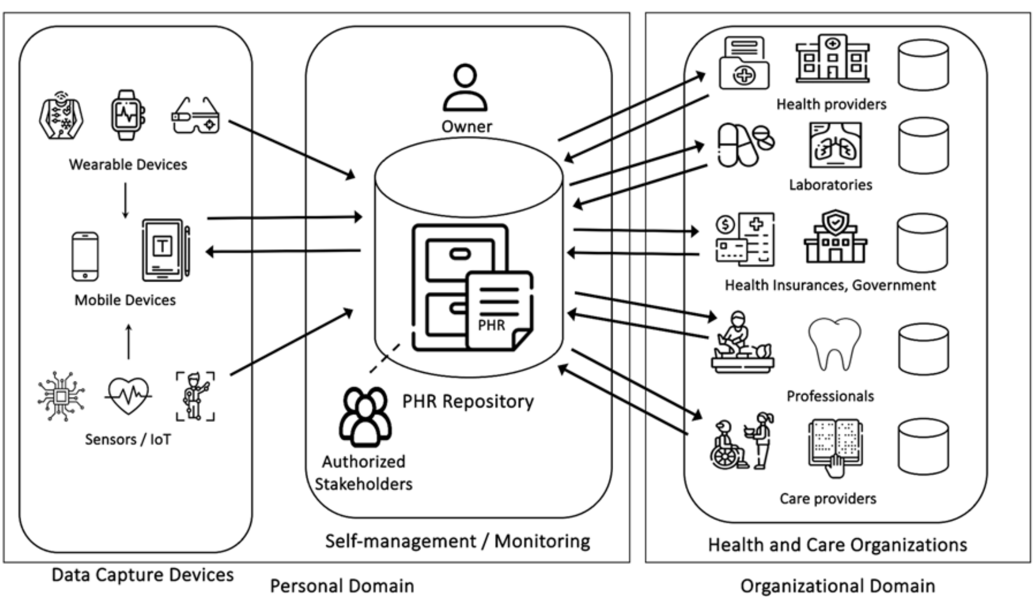
The introduction of the individual as the main actor in information management implies design differences between PHR-based and EHR-based applications. The needs that people want to satisfy in PHR applications have been identified, highlighting the following (Fuji et al., 2012):

- Sharing health information with others, receiving feedback based on entered health information.
- Present information in simple terms.
- Ensuring the security and privacy of health information.
- Opportunity to communicate with healthcare providers.
- Interoperability with the providers'-based record.
- Generate printed reports of health information.
- Ability to create new sections in the PHR to store additional information from other medical areas.
- Customize the user interface.
- Access control to third parties to display subsets of health information.
- Having personalized support in the PHR based on an individual's abilities and needs.

The most relevant concerns and challenges of PHR applications have also been identified, with the appropriate management of privacy being the most relevant one (Jacquemard et al., 2020). Privacy management in this context is highly complex because it is not only a technological challenge but also has a strong ethical and legal component. Along with privacy, other concepts such as consent and privacy are linked. Following the logic of PHRs, individuals must have the ability to share all or a subset of their information with third parties. Based on this exchange of information, three main types of PHR applications have been cataloged: provider-tethered, payor-tethered, and standalone (Vincent et al., 2008). The latter, as they are not linked to any context, are ideal for communication with systems from different domains and it's therefore easier to store heterogeneous information in them.

Figure 1 shows the data flow between different EHRs and personal IoT sensors in a PHR environment.

Figure 1. General diagram of relationships between PHR and HER



Everyone must have a unique PHR that can communicate with several EHRs in a bidirectional way to read or add information. The PHR's owner will have the ability to allow access to external domain systems or third parties. In the personal domain, PHRs are responsible for storing the information generated by different sensors interacting with mobile devices.

Privacy

In the United States, the 1996 HIPAA law[4] dictates how to adequately protect medical data, but since it predates the digital revolution, its spectrum of action has had to be expanded with the drafting of the GINA law[5]. The European Union has also developed legislation to protect the privacy of individuals' medical information (Council regulation (EU) no 2016/679, 2016) and even the ISO has developed standards for this purpose (ISO 27799, 2016). The defense of the right to privacy can conflict with other rights such as autonomy, which can be defined as a person's right to make choices and views, and to take actions based on personal values and beliefs. The goal is to promote the independence of people and not take any actions that would increase their dependence on third parties.

Achieving a balance between these concepts is an enormously complex challenge. To handle privacy of information issues, we must apply security measures. Although (Fernández-Alemán et al., 2013) focus privacy issues in EHR systems on Confidentiality, Integrity, and Availability (CIA), with the proliferation of cloud systems in recent times, these requirements have had to be adjusted to new architectures (Sahi et al., 2021). An additional challenge in PHR systems is that health providers are not the only stakeholder that has control over patients' information. PHR systems must provide individuals with the necessary tools to manage their information with third parties.

To solve the issues described above, we propose the use of a decentralized architecture that combines the SOLID paradigm (Social Linked Data) (Sambra et al., 2016), semantic web technologies, along ontologies, and medical data standards. Its main feature is the storage of dental information in SOLID Personal Online Datastores (PODs), i.e., a single repository under the control of the patient and independent of any medical institution.

This architecture will have a layer composed of heterogeneous computational agents—a computational agent, in the context of this architecture, refers to a software program or system in a personal or organizational domain that is capable of communicating bidirectionally with the personal repositories, processing the data stored there and adapt them to its domain—independent of any stakeholder in charge of the exchange of information between organizations and patients. It will be able to recover, validate, and transparently transform the data while respecting the right of consent for the use of patient information. Besides, the information storage will follow the FHIR standard along with Shape Expressions (ShEx), allowing any computational agent to properly consume the information. FHIR archetypes will be used to represent and share information in a homogeneous yet flexible manner. FHIR has garnered extensive recognition and adoption across diverse domains, including the fields of medicine and medical research (Vorisek et al., 2022), owing to its notable attributes such as modularity and scalability. ShEx, on the other hand, is used to describe data structures for querying, requesting or publishing information. Finally, SOLID PODs will be used to store the data in a decentralized but interconnected manner that allows users to keep the ownership of their personal data.

The granularity of this architecture allows small businesses to exchange clinical information with their patients without the need for expensive developments.

This approach presents several challenges that must be addressed in the future, including the following: 1) handling the computational cost in the consumption of stored information; 2) developing a new business layer that deals with the hosting of the repositories; and 3) promoting the acceptance of the proposed paradigm shift by organizations and patients.

STATE OF THE ART

Security Concerns

The kind of data handled in health-related applications requires special care regarding security concerns. Specifically, in this subsection, we will review state-of-the-art work related to access control, role systems, anonymization, and encryption techniques. Access control is the tool with which access to information can be modeled by creating a role policy based on the user. User profiles can be created and granular access to information can be provided. The combination of these tools makes it possible to manage consent and establish how users can manage the information and to what extent. Encryption is a critical tool in data security. If a system is the victim of a cyberattack, we can prevent the information from being read or modified by hackers.

Currently, there are two main courses of action to achieve encryption. The first one is to use some encryption algorithm. The most used encryption systems are public key encryption (PKE) (Chenthara et al., 2019), and symmetric key encryption (SKE). More sophisticated algorithms are also used, including attribute-based encryption (ABE) (Fabian et al., 2015), fully homomorphic encryption (FHE) (Kocabas & Soyata, 2015; Page et al., 2014), searchable encryption (SE) (Liu et al., 2016; Yang & Ma, 2016), and five-dimensional hyper-chaotic map (FDHC) to protect the transmission of biomedical images over IoT networks (Kaur et al., 2021). However, such approaches require higher computational costs compared to PKE and SKE.

The second alternative to achieve encryption is using blockchain technology. The main advantages of such an approach are the complexity of encryption that it can provide, and its potential to ensure data integrity and data decentralization. However, blockchain-based approaches must deal with issues related to high computational (Tandon et al., 2020) and monetary cost (Krause & Tolaymat, 2018; Stoll et al., 2019), that prevent their use in projects of a certain size.

To overcome these limitations, more sophisticated solutions are being explored. Their objective is to reduce the computational load of the entities involved in the system. One line of research is to reduce the associated cost of a fine-grained searchable encryption scheme with the help of blockchain technology (Mamta et al., 2021).

Interoperability

The IEEE defines interoperability as "The ability of two or more systems or components to exchange information and to use the information that has been exchanged" («IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries», 1991). Sheth (1999) expanded this concept on three main levels: syntactic, structural, and semantic; the latter being the most challenging one, as it may require being able to process information produced or handled by different computational agents.

The goal of structural interoperability is for the information transmitted to be understandable by a domain expert. This can be achieved by combining the use of international clinical standards with knowledge representation using Resource Description Format (RDF). As it has been proclaimed in the Yosemite Manifesto, RDF is "the best available candidate for a universal healthcare Exchange language" (Yosemite Manifesto, 2013). Additionally, RDF can be used as a knowledge representation language enabling reasoning capabilities that could be extended to support different types of reasoning, like fuzzy or defeasible[6] (Salloum & Tekli, 2021). This gives it great flexibility in the number of uses it can provide.

These standards were created from the so-called archetype model, which consists of the definition of clinical concepts using artifacts called archetypes that allow for combining and restricting the entities of a reference model to define clinical concepts. Some widely disseminated standards are FHIR, developed by HL7 (FHIR, 2023), ISO-13606 (ISO 13606, 2019), and ISO-20514 (ISO 20514, 2005), whose main goal is the homogenization of the data exchange in medical environments. This is achieved by creating abstract concept structures from sets of individual parameters such as weight, height, or blood pressure, so they can be reused in different systems (Maldonado et al., 2020). Each parameter contains all the necessary concepts to be self-explanatory. They work as a system of independent blocks that can be combined in such a way that they can represent large abstract entities such as medical records (Mandel et al., 2016). It is also necessary to use terminologies that consistently represent clinical information, favoring its reuse in different institutions and allowing the efficient retrieval and reuse of information based on its meaning. Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) is a widely accepted terminology made up of hierarchically organized concepts, descriptions, and relationships that allow any type of clinical information to be represented at any level of detail (Systemized Nomenclature of Medicine - Clinical Terminology (SNOMED-CT), 2023). The combination of standards and terminologies improves the retrieval and reuse of clinical information, taking a further step towards achieving semantic interoperability.

The purpose of semantic interoperability is for an information system to be able to understand the meaning of the information it is handling. To achieve this, it is necessary to provide the information with the necessary context to describe its meaning in the modeled domain. Through the combination of metadata and ontologies, it is possible to build a model that allows the addition of the necessary conceptualization for a correct interpretation of the data.

Metadata can be defined as data about data. Metadata allows you to describe domain information about the underlying data and represent details about the format of the data or explain constraints. Metadata describing the semantics of the modeled domain can be added by a separate process known as semantic annotation, which can be done manually (Kiryakov et al., 2004), semi-automatically, or automatically (Oren et al., 2006).

One of the biggest challenges of semantic interoperability is getting the different stakeholders involved to agree on the use of existing ontologies to define a common framework in which information will be managed. In a decentralized context, this is hard to achieve due to the heterogeneous nature of the information stored.

PROPOSAL

We propose an architecture that achieves the interoperability of dental data from heterogeneous data sources by applying the SOLID principles (Social Linked data), RDF and REDF-related technologies (including ShEx), and the concept of POD in conjunction with the PHR paradigm.

With this approach, each patient will have full control over their unified dental information, it will be possible to integrate EHR data coming from different sources, and the information will be stored and served in a way that each organization is able to take part in its context. The data will be stored in PODs owned by the patient in a decentralized system. Thanks to those PODs, the patient will have the ability to accurately manage access control to that information by third parties.

SOLID PODs are based on existing W3C standards that enable the authentication process based on WebID and Web Access Control. These features allow all the information of a given patient to be stored in a single repository and establish mechanisms that allow the bidirectional exchange of information.

SOLID applications rely on WebID to provide a decentralized authentication. WebIDs belong to a global ID space which has the form of an HTTP(S) URI and global single sign-on to the users. Every user must be registered with an identity provider that stores the user's WebID profile document associated with a cryptographic key. The general idea behind WebID is to allow different computational agents (people, organizations, etc.) to link their data and share information, creating a network of trust between peers.

Each resource stored in a POD has its unique URI, too. With this very precise identification of users and resources, SOLID can use Access Control Lists (ACLs) to establish a granular level of access to resources stored as a file system managed by the POD owner.

An ecosystem in which two different types of domains coexist must be developed. On the one hand is a personal domain whose purpose is focused on the development of tools that give the owner of the data full control over them. On the other hand, an organizational domain is aimed at allowing the different medical providers or physicians to interact with the patient's personal data repository. The applications of both domains have the following points in common: they work with the personal data source independently, they define an output model that transforms the input data so that it can be used for a specific purpose, and they can add new information to the personal repository so it can be processed by other applications.

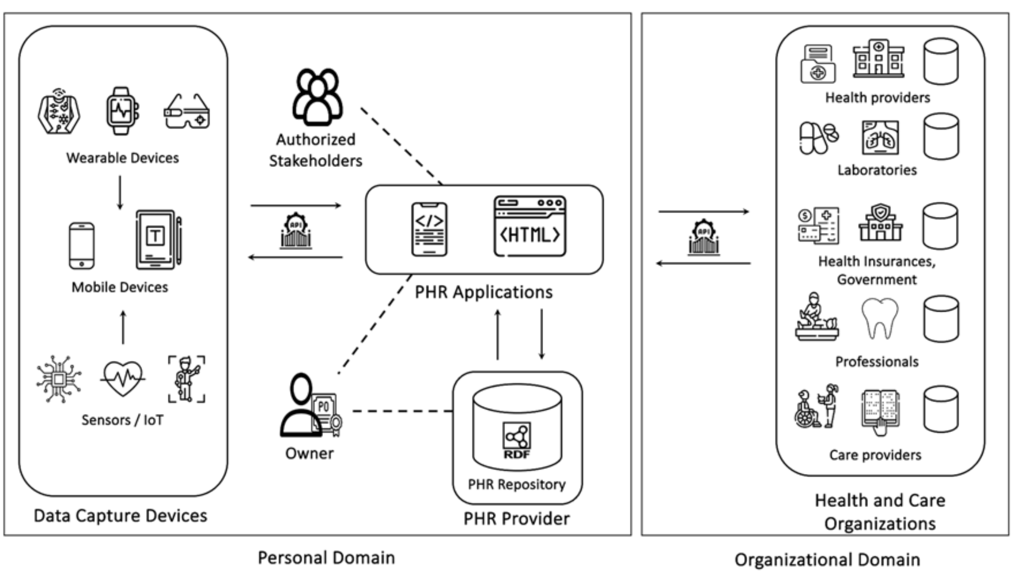
In the personal domain, applications have the main goal of facilitating the owner's personal use of information, managing such information, and handling data access to control what third parties can query/modify what part of the information. Meanwhile, organizational field applications aim at integrating information between PODs and the organizations' systems. Figure 2 shows a main diagram of the proposed architecture.

Storing the patient's clinical data in their personal repository in RDF format favors its interoperability, as it facilitates the unification between heterogeneous data sources thanks to its simple structure that allows data, metadata, terminology, and ontologies to be represented in a unified way. RDF makes it possible for information from disparate sources and structures to be combined into a single federated collection of Linked Data. RDF is designed for "distributed extensibility," which allows the consumer to select understandable information without needing to fully understand it.

The flexibility of RDF requires the implementation of a mechanism that allows each system to retrieve and adapt the information stored in the patient's repository according to their needs. For this, schemas must be defined that serve as patterns in a transformation process. These templates must be applied between the input and output models defining the archetypes used in the modeling of data.

ShEx is a language that allows to modeling graphs (sets of RDF triples) in a precise and formal way. By defining a ShEx schema, one defines the requirements that a graph must satisfy (Prud'hommeaux et al., 2014). If the graph satisfies the requirements defined in the schema, it means that the set of triples is valid—it conforms to the constraints described in the schema. These schemas can represent a part of an entity of the output model or even its entirety, which validates the correspondences between the

Figure 2. General architecture diagram with decoupling between the PHR repository, services, and organizational domains



input and output models. They also allow an inference testing service that uses the given schema on the input graph and applies a reasoner capable of inferring the necessary transformations that permit adapting the information to the output model by adding new triples and performing transformations. Shape Expressions have demonstrated its efficiency as a tool to harmonize differences between heterogeneous data in different areas, allowing information to be transmitted unambiguously by finding inconsistencies or errors in data models (Candela, 2023; Candela et al., 2022; Thornton et al., 2019; Thuluva et al., 2018). For all these reasons, Shape Expressions have proven to be the best candidate for FHIR archetype validation (Solbrig et al., 2017).

The use of archetypes is a very powerful resource that is not limited to the structuring of clinical information. It also allows inserting additional data that enriches the context by eliminating possible ambiguities or adding additional structures that can define concepts. An example of such new non-clinical information would be the authorship of the different entries of the stored clinical information, which can be useful for obtaining the traceability of the different health providers, clinics, or professionals that have interacted with the author in the past.

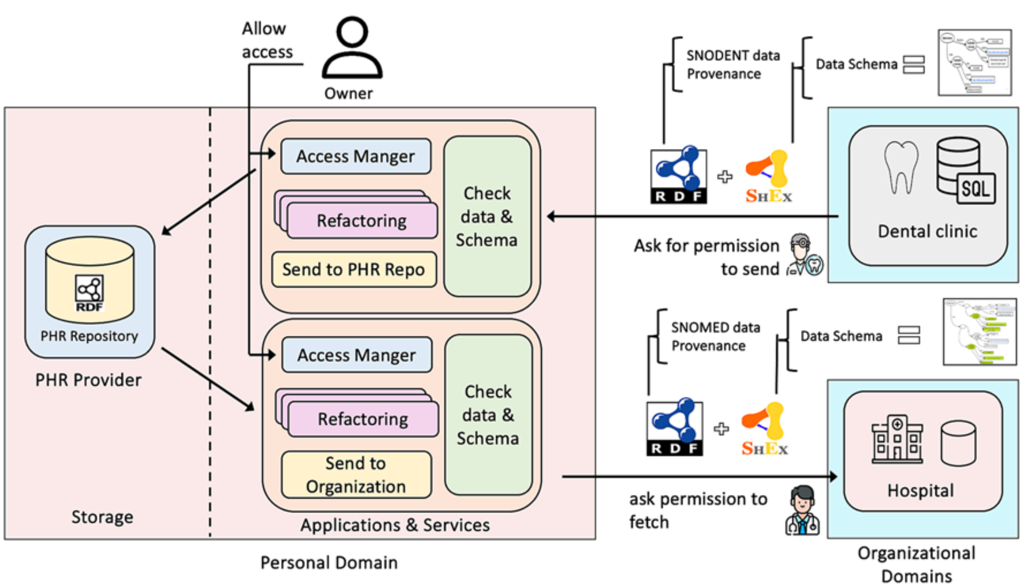
Figure 3 shows the information exchanged between the personal and organizational domains, as well as a main approach to the services that independent agents should provide.

Use Case

The technological approach described in the previous section will be exposed with the following use case.

Alice has been suffering from acute tooth pain for a couple of days, so she decides to go to Bob's dental clinic. After an examination, Bob discovers that the cause of the pain is an infection in the lower left second molar. When the appointment finishes, Bob stores the appointment information on his computer system. Bob uses a very simple computer application designed for small dental clinics which uses the SNODENT terminology (Systematized Nomenclature of Dentistry (SNODENT), 2022). Bob needs to add to his system only three fields:

Figure 3. Data flow and services are provided by independent agents



- The identifier of the affected tooth.
- The identifier of the service performed.
- The date on which the intervention took place.

Then, Alice asks Bob to send the procedure information to her PHR repository. Bob agrees and, using a simple graphical interface on his system, initiates an independently developed workflow whose outputs are an RDF graph that contains the intervention information together with the data that identifies Bob as the author of the procedure. The result also includes a ShEx schema that describes and validates the generated data graph. Figure 4 shows the data and ShEx content sent to System_A by Bob.

Alice tells Bob the URL of the service that will handle the information storage process. Bob accesses it through his browser, uploads the files in a web form, and enters Alice's email.

System_A initializes a workflow that compares the schema with the RDF data and checks that the data conforms to it. The system then sends an email to Alice with a preview of the information, asking if she gives System _A permission to store the data in her PHR repository.

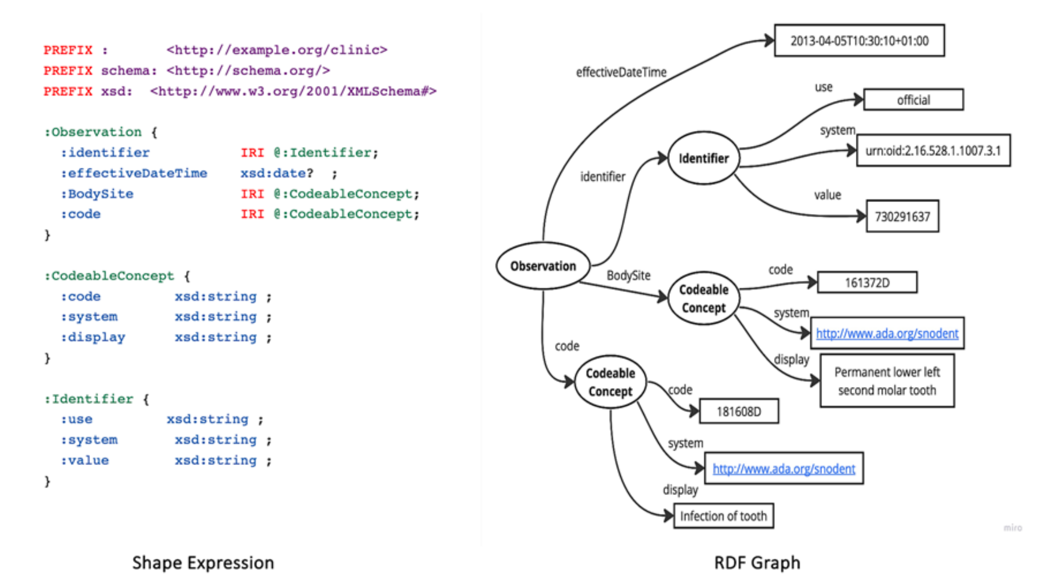
Alice notifies System_A of the acceptance of the storage request. To do so, she sends the URL of her PHR repository in response to the request. When System_A requests to communicate with the repository, the latter checks that it has written permission on its ACL.

Charlie is Alice's family doctor, who informs her during her routine check-up that the hospital's computer system can store the dental records of patients. He insists that it would be convenient to have access to such information, even if it was not produced by Charlie's hospital.

Dental information could be entered into the system manually. To do so, Charlie would have to ask Alice a set of questions whose answers will serve to complete the necessary records in the system. However, another possibility would be to integrate the dental information from Alice's PHR repository into the hospital's system. Alice chooses the second option. For this, she needs to grant read permissions to her PHR repository so that System_B can query its data.

Charlie uses a graphical interface based on vocabulary terms known by the doctor. Such an interface produces a ShEx schema describing the structure of the target information. The schema is based on FHIR archetypes and uses SNODENT terminology. Such schema is sent to System_B.

Figure 4. Data and shape expression sent to System_A by Bob



Alice allows System_B to access her PHR repository, just like System_A did. At that time, the system walks through the repository comparing the ShEx schema with the stored graphs. The triples that match the schema are returned in a single file.

Charlie loads the file generated in the hospital's computer system. The IT team developed a workflow that transforms the SNODENT terminology triplets to SNOMED, which allows Alice's clinical history to be automatically loaded into the hospital's system.

Alternatively, Alice's dental information may not need to be stored in the hospital's computer system but queried and consumed directly from Alice's POD using the described workflow whenever such information is needed by the hospital.

DISCUSSION

This paper presents a technical approach that allows the semantic interoperability of dental data between independent systems belonging to different clinical areas, without the need for the data exchanged to have a common structure or terminology. To this end, the clinical information exchanged must be serialized in RDF, following archetypes adjusted to the FHIR standard, and sent together with ShEx schemas that define the structure of the data. The joint sending of the information and the schemas allows any computational agent to receive the data, validate it, and make the necessary modifications to enrich or adapt it for consumption by an external system.

The autonomy that the computational agents acquire in this approach enables the independence of the services with respect to the repositories. It allows the source of information to be decoupled from any agent designed to process it, opening the door to the development of decentralized architectures in which the owner of the information has full control over the repository and can grant permissions to the agents that require such information. With this, the data owners have extensive control over their personal dental information. This promotes information exchange with external systems transparently, overcoming restrictions related to the structure or terminology used, and without the need to establish a prior agreement in which common ontologies and vocabularies must be used in the exchange of information. Furthermore, the decoupling between the PHR repository and the services empowers the data owner to explicitly distinguish between two purposes for the systems. The first one is the correct management of the hosting of the repositories. The second one is to establish mechanisms that enrich the information they receive as input. The organizational domains will be able to consume external information by carrying out a granular development that adjusts to their needs, developing workflows that do not affect the rest of the computer system. Alternatively, organizations could use intermediate services that help the intercommunication between the PHR repository and their system.

In addition, querying techniques across decentralized environments (Taelman & Verborgh, 2023) could be used by medical institutions to improve disease diagnosis, facilitating the creation of more complex and heterogeneous datasets that help develop more precise machine learning or deep learning techniques (Shankar et al., 2022).

The following challenges posed by this architecture need to be addressed. First, the computational cost that some agents performing transformation or inference operations may demand could cause a performance bottleneck making the service useless. Also, the use of ShEx can be an entry barrier in the development of domain-independent applications. It is also possible that an agent company gains a dominant position in the market, which would lead to a limitation in the use of vocabularies.

The development of workflows that integrate information from PHR repositories in an organizational domain requires expert knowledge of ShEx, FHIR, as well as the output system. However, this barrier could be mitigated by developing general tools based on graphical interfaces that are easy to use for practitioners and built on top of a controlled set of vocabularies. Such software could be reused by different dental institutions so the live input produced by different practitioners could easily be integrated into PODs and the PODs' content could be easily visualized by practitioners. Such software would be able to automatically produce the necessary vocabulary mappings, FHIR, and

ShEx content based on the user's actions, and therefore it would hide the complexity of the information exchange process. However, this kind of software reduces the gap between existing systems and our proposal, but it would not be enough to synchronize each institution's backend with our system. For such a goal, ad-hoc developments may be required.

PHR repository hosting services must face new types of attacks and must invest resources in developing the necessary security measures. The acceptance of this model by end users (patients and professionals) can mean a very strong paradigm shift that causes rejection.

Our proposal should also address potential performance issues at different stages of the information exchange. The performance of the different computational agents that interact in this approach must be evaluated, as it may affect the performance of the general process when dealing with outdated systems/hardware or when dealing with too much information. The same goes for the computational cost needed by the validation processes that require the data sent and the schemas, or the inference processes carried out by the independent agents. The ability to integrate dental information with clinical information from other areas should be studied as well. Finally, accepting the role change of people who now have an active role in managing your data is still a work in progress that must be evaluated in real cases under different circumstances.

CONCLUSION

This paper highlights the need for an ecosystem that allows for the interoperability of PHRs between different systems independent from each other. A technical solution that combines FHIR archetypes defined by stakeholders with Shape Expressions in a decentralized environment—in which there is a decoupling between the personal information repository and any service that wishes to consume the data—is proposed. This solution allows patients to unify their dispersed dental information between different systems, as well as grants them the ability to share it with other stakeholders in full or in part, overcoming the limitations derived from the structure and terminology of each domain. This solution allows all stakeholders to validate and verify the authorship of any piece of information transparently.

The bidirectional data exchange between individuals' Personal Online Datastores (PODs) and healthcare provider systems not only empowers individuals to centralize and uphold their medical information but also streamlines healthcare providers' access to and utilization of this data overcoming the challenges of information silos.

The theoretical proposal described in this paper should be evaluated in future work by developing a prototype that allows dental information to be exchanged between practitioners and patients. We will test such prototype in different scenarios, aiming to detect and tackle potential challenges that may arise when deploying a system based on our proposals.

ACKNOWLEDGMENT

This project has been partially funded by the Spanish Ministry of Science and Innovation project: Applying Knowledge Graphs to research data interoperabiLIty, PID2020-117912RB-C21.

REFERENCES

Candela, G. (2023). An automatic data quality approach to assess semantic data from cultural heritage institutions. *Journal of the Association for Information Science and Technology*, *74*(7), 866–878. doi:10.1002/asi.24761

Candela, G., Escobar, P., Sáez, M. D., & Marco-Such, M. (2022). A Shape Expression approach for assessing the quality of Linked Open Data in libraries. *Semantic Web*, *14*(2), 159–179. doi:10.3233/SW-210441

Chenthara, S., Ahmed, K., Wang, H., & Whittaker, F. (2019). Security and privacy—Preserving challenges of e-Health solutions in Cloud computing. *IEEE Access : Practical Innovations, Open Solutions*, *7*, 74361–74382. doi:10.1109/ACCESS.2019.2919982

Chevreul, K., Berg Brigham, K., Durand-zalesk, I., & Hernández-Quevedo, C. (2015). France. *Health Systems Review*, *17*, 3. PMID:26766545

Connecting for Health. The personal health working group final report. (2003). Markle Foundation. https://markle.org/publications/1429-personal-health-working-group-final-report/

Council regulation (EU) no 2016/679. (2016). https://data.europa.eu/eli/reg/2016/679/oj

Fabian, B., Ermakova, T., & Junghanns, P. (2015). Collaborative and secure sharing of healthcare data in multi-clouds. *Information Systems*, *48*, 132–150. doi:10.1016/j.is.2014.05.004

Fellows, J. L., Atchison, K. A., Chaffin, J., Chávez, E. M., & Tinanoff, N. (2022). Oral health in America. *The Journal of the American Dental Association*, *153*(7), 601–609. doi:10.1016/j.adaj.2022.04.002 PMID:35643534

Fernández-Alemán, J. L., Señor, I. C., Lozoya, P. Á. O., & Toval, A. (2013). Security and privacy in electronic health records: A systematic literature review. *Journal of Biomedical Informatics*, *46*(3), 541–562. doi:10.1016/j.jbi.2012.12.003 PMID:23305810

FHIR (5.0.0). (2023). http://hl7.org/fhir/

Fuji, K. T., Abbott, A. A., Galt, K. A., Drincic, A., Kraft, M., & Kasha, T. (2012). Standalone personal health records in the United States: Meeting patient desires. *Health and Technology*, *2*(3), 197–205. doi:10.1007/s12553-012-0028-1

Gupta, B. B., Li, K. C., Leung, V. C., Psannis, K. E., & Yamaguchi, S. (2021). Blockchain-assisted secure fine-grained searchable encryption for a cloud-based healthcare cyber-physical system. *IEEE/CAA Journal of Automatica Sinica, 8*(12), 1877-1890. 10.1109/JAS.2021.1004003

IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. (1991). *IEEE Std 610*. 10.1109/IEEESTD.1991.106963

ISO 13606. (2019). [Standard].

ISO 20514. (2005). [Standard].

ISO 27799. (2016). [Standard].

Jacquemard, T., Doherty, C. P., & Fitzsimons, M. B. (2020). Examination and diagnosis of electronic patient records and their associated ethics: A scoping literature review. *BMC Medical Ethics*, *21*(1), 76. doi:10.1186/s12910-020-00514-1 PMID:32831076

Kaur, M., Singh, D., Kumar, V., Gupta, B. B., & Abd El-Latif, A. A. (2021). Secure and energy efficient-based e-Health care framework for green internet of things. *IEEE Transactions on Green Communications and Networking*, *5*(3), 1223–1231. doi:10.1109/TGCN.2021.3081616

Kiryakov, A., Popov, B., Terziev, I., Manov, D., & Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. *Journal of Web Semantics*, *2*(1), 49–79. doi:10.1016/j.websem.2004.07.005

Kocabas, O., & Soyata, T. (2015). Utilizing homomorphic encryption to implement secure and private medical cloud computing. *2015 IEEE 8th International Conference on Cloud Computing*, 540-547. doi:10.1109/CLOUD.2015.78

Krause, M. J., & Tolaymat, T. (2018). Quantification of energy and carbon costs for mining cryptocurrencies. *Nature Sustainability*, *1*(11), 711–718. doi:10.1038/s41893-018-0152-7

Lee, P. C., Wang, J. T. H., Chen, T. Y., & Peng, C. H. (2022). *Digital Health Care in Taiwan: Innovations of National Health Insurance*. Springer International Publishing. doi:10.1007/978-3-031-05160-9

Liu, Z., Weng, J., Li, J., Yang, J., Fu, C., & Jia, C. (2016). Cloud-based electronic health record system supporting fuzzy keyword search. *Soft Computing*, *20*(8), 3243–3255. doi:10.1007/s00500-015-1699-0

Maldonado, J. A., Marcos, M., Fernández-Breis, J. T., Giménez-Solano, V. M., Legaz-García, M. del C., & Martínez-Salvador, B. (2020). Clin-Ik-Links: A platform for the design and execution of clinical data transformation and reasoning workflows. *Computer Methods and Programs in Biomedicine*, *197*, 105616. doi:10.1016/j.cmpb.2020.105616 PMID:32629294

Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association : JAMIA*, *23*(5), 899–908. doi:10.1093/jamia/ocv189 PMID:26911829

Oren, E., Moller, K. H., Scerri, S., Handschuh, S., & Sintek, M. (2006). What are Semantic Annotations? Relatório técnico. *DERI Galway*, *9*, 62.

Page, A., Kocabas, O., Ames, S., Venkitasubramaniam, M., & Soyata, T. (2014). Cloud-based secure health monitoring: Optimizing fully-homomorphic encryption for streaming algorithms. *2014 IEEE Globecom Workshops (GC Wkshps),* 48-52. 10.1109/GLOCOMW.2014.7063384

Prud'hommeaux, E., Labra Gayo, J. E., & Solbrig, H. (2014). Shape expressions: An RDF validation and transformation language. *Proceedings of the 10th International Conference on Semantic Systems*, 32-40. doi:10.1145/2660517.2660523

Sahi, A., Lai, D., & Li, Y. (2021). A review of the state of the art in privacy and security in the eHealth cloud. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 104127–104141. doi:10.1109/ACCESS.2021.3098708

Salloum, G., & Tekli, J. (2021). Automated and personalized nutrition health assessment, recommendation, and progress evaluation using fuzzy reasoning. *International Journal of Human-Computer Studies*, *151*, 102610. doi:10.1016/j.ijhcs.2021.102610

Sambra, A. V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., Zagidulin, D., Aboulnaga, A., & Berners-Lee, T. (2016). *Solid: A platform for decentralized social applications based on linked data*. Academic Press.

Shankar, K., Perumal, E., Mohamed, Taher, F., El-Latif, B. B. & Abd, A. A. (2022). *Synergic deep learning for smart health diagnosis of COVID-19 for connected living and smart cities*. Association for Computing Machinery.

Sheth, A. P. (1999). Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In *Interoperating geographic information systems* (pp. 5–29). Springer US. doi:10.1007/978-1-4615-5189-8_2

Solbrig, H. R., Prud'hommeaux, E., Grieve, G., McKenzie, L., Mandel, J. C., Sharma, D. K., & Jiang, G. (2017). Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx). *Journal of Biomedical Informatics*, *67*, 90–100. doi:10.1016/j.jbi.2017.02.009 PMID:28213144

Stoll, C., Klaaßen, L., & Gallersdörfer, U. (2019). The carbon footprint of Bitcoin. *Joule*, *3*(7), 1647–1661. doi:10.1016/j.joule.2019.05.012

Systematized Nomenclature of Dentistry (SNODENT) (2.0.0). (2022). https://www.ada.org/resources/practice/dental-standards/snodent

Systemized Nomenclature of Medicine—Clinical Terminology (SNOMED-CT) (Versión 20230901). (2023). https://www.nlm.nih.gov/healthit/snomedct/international.html

Taelman, R., & Verborgh, R. (2023). *Evaluation of Link Traversal Query Execution over decentralized environments with structural assumptions.* arXiv preprint arXiv:2302.06933.

Tandon, A., Dhir, A., Islam, A. K. M. N., & Mäntymäki, M. (2020). Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda. *Computers in Industry*, *122*, 103290. doi:10.1016/j.compind.2020.103290

Thornton, K., Solbrig, H., Stupp, G. S., Labra Gayo, J. E., Mietchen, D., Prud'Hommeaux, E., & Waagmeester, A. (2019). Using shape expressions (ShEx) to share RDF data models and to guide curation with rigorous validation. In *The Semantic Web: 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2–6, 2019, Proceedings 16* (pp. 606-620). Springer International Publishing. doi:10.1007/978-3-030-21348-0_39

Thuluva, A. S., Anicic, D., & Rudolph, S. (2018). IoT semantic interoperability with device description shapes. In *The Semantic Web: ESWC 2018 Satellite Events: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers 15* (pp. 409-422). Springer International Publishing. doi:10.1007/978-3-319-98192-5_56

Torres-Urquidy, M. H., Powell, V., Geist, S. M. R. Y., Mishra, S., Chaudhari, M., & Allen, M. (2019). Health information technology considerations of medical and dental data integration. *Integration of Medical and Dental Care and Patient Data*, 155-206. 10.1007/978-3-319-98298-4_11

Vincent, A., Kaelber, D. C., Pan, E., Shah, S., Johnston, D., & Middleton, B. (2008). A patient-centric taxonomy for personal health records (PHRs). *AMIA ... Annual Symposium Proceedings - AMIA Symposium. AMIA Symposium*, *2008*, 763. PMID:18998912

Vorisek, C. N., Lehne, M., Klopfenstein, S. A. I., Mayer, P. J., Bartschke, A., Haese, T., & Thun, S. (2022). Fast healthcare interoperability resources (FHIR) for interoperability in health research: Systematic review. *JMIR Medical Informatics*, *10*(7), e35724. doi:10.2196/35724 PMID:35852842

Vujicic, M. (2022). Time for dental care to be considered essential in US health care policy. *AMA Journal of Ethics*, *24*(1), E57–E63. doi:10.1001/amajethics.2022.57 PMID:35133729

Yang, Y., & Ma, M. (2016). Conjunctive keyword search with designated tester and timing enabled proxy re-encryption function for e-health clouds. *IEEE Transactions on Information Forensics and Security*, *11*(4), 746–759. doi:10.1109/TIFS.2015.2509912

Yeh, M. J., & Saltman, R. B. (2019). Creating online personal medical accounts: Recent experience in two developed countries. *Health Policy and Technology*, *8*(2), 171–178. doi:10.1016/j.hlpt.2019.05.004

Yosemite Manifesto. (2013). Retrieved from http://www.yosemitemanifesto.org/

ENDNOTES

[1] https://scitechdaily.com/obama-administration-pumped-27-billion-into-electronic-health-records-doctors-give-an-f/?utm_content=cmp-true

[2] https://digital-strategy.ec.europa.eu/en/policies/electronic-health-records

[3] https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf

[4] https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf

[5] https://www.govinfo.gov/content/pkg/PLAW-110publ233/pdf/PLAW-110publ233.pdf

[6] https://ercim-news.ercim.eu/en131/r-i/plausible-reasoning-that-mimics-human-argumentation#:~:text=A%20convenient%20way%20to%20express,%2C%20relationships%2C%20dependencies%20and%20implications

Hugo Lebredo is currently pursuing his doctoral studies while working as a lecturer at the University of Oviedo, Spain. Since 2019, he has been an active member of the WESO (Web Semantics Oviedo) Research Group, under the leadership of Jose Emilio Labra Gayo. His research specialization revolves around the semantic web, with a particular emphasis on improving the interoperability of Electronic Health Records across various domains.

Daniel Fernández-Alvarez is a PhD in Computer Science and Associate Professor at the University of Oviedo, Spain. He has been a member of the WESO (Web Semantics Oviedo) Research Group, led by Jose Emilio Labra Gayo, since 2014. He specializes in topics related to semantic web, and, specifically, RDF shapes. Despite he has worked in different fields, such as record linkage, data analysis, and data normalization, most of his work is focused on the automatic extraction of RDF shapes from knowledge graphs/natural language. He is a member of the ShEx IEEE group and the W3C's Shape Expressions Community Group.

Jose Emilio Labra-Gayo is a Full Professor (Catedrático) in University of Oviedo, Spain. He is the founder and main researcher of WESO (Web Semantics Oviedo) research group, which collaborates with different companies around the world applying semantic web technologies to solve practical problems. Member of Data Shapes W3C Working Groups and of Shape Expressions, SHACL, RAX (RDF and XML interoperability) and Best practices of Multilingual Linked Open Data W3C community groups. Previously, coordinator of the Master in Web Engineering - University of Oviedo (2016-2017), and Dean of the School of Computer Science Engineering - University of Oviedo (2004-2012).