APPLICATION OF SEMANTIC INTEGRATION METHODS FOR CROSS-AGENCY INFORMATION SHARING IN HEALTHCARE

Yu. M. Akatkin, E.D. Yasinovskaya, M.G. Bich, A.V. Shilin

Plekhanov Russian University of Economics, 36 Stremyanny per., Moscow, 117997, Russia

E-mail: u.akatkin@semanticpro.org, elena@semanticpro.org, misha@semanticpro.org, a.shilin@semanticpro.org

Cross-agency Information sharing is a basic feature of digital economy in public sector. Heterogeneous environment is inherent for the public sector as well as for other industries, which are ready to digital transformations. The application of integration methods should guarantee the achievement of unambiguous meaningful interpretation of data. The article represents comparative analysis of basic integration approaches: (1) classic interaction via mediators, (2) integrated data model (XML-based models implementation) and (3) semantic integration. We defined the advantages of semantic integration and confirmed the possibility to use semantic core as a basis for a digital health ecosystem. We implemented the example of semantic integration in healthcare within the project of Plekhanov Russian University of Economics named the "Center of semantic integration".

Keywords: semantic interoperability, semantic integration, cross-agency interaction, information sharing, domain data model, information exchange, digital ecosystem, digital health, digital economy

1. Introduction

Digitalization is a priority for the future economic development of the Russian Federation. Digital Health is one of the main industry directions that expands the federal program "Digital Economy of the Russian Federation"[1]. The possibility to create digital health services in compliance with the principles of semantic interoperability in a heterogeneous information environment taking into account a multitude of actors (medical organizations, patients, doctors, insurance companies, developers, etc.) and a multitude of diverse information resources is a scientific and practical task, which is important for digital transformation in healthcare.

Currently, researchers as well as technical community around the world are working on data standardization, semantic assets' cataloguing and reuse, together with data models design and dissemination, unification integration methods in order to implement them for the information sharing in various knowledge domains.

At the semantic level, a unified description of data supports the integration, taking into account the semantic properties in the context of a single ontology of the subject area [1]. A number of works, for example [1-6], describes semantic integration methods and their advantages over the traditional approaches. Despite the fact that "semantic modeling has been the subject of research since the late 1970s" [3], the problem of heterogeneous data semantics that arise both within information systems and during information exchange between them remains relevant [4].

The development of methods for the practical use of semantic integration is a key enabler for digital transformation both at the national and sectoral level.

The objective of this article is to identify the advantages of semantic integration for the implementation of information sharing principles, essential to the building of a digital health ecosystem [7, 8]. To achieve this aim we made a comparative analysis of the existing approaches used for the integration of data in information systems [6]. We confirmed the results of the analysis by the approbation of the chosen method on real services of cross-sector interaction between the information systems of the Ministry of Health of Russia and the Social Insurance Fund.

2. Comparative analysis of the existing data integration approaches

The researchers identify three approaches to the data integration of information systems in a heterogeneous environment: (1) classic interaction via mediators, (2) integrated data model and (3) semantic integration [6]. Russian practice of cross-agency interaction shows the frequent use of the first method. NIEM (USA) [9, 10] is an example of the implementation of integrated XML data models in the public sector and the European Union actively promotes the use of semantic integration within ISA[2] program, providing interoperability solutions for public administrations, businesses and citizens [11].

To analyze the advantages or disadvantages of the existing integration methods, we consider the possibility to implement them in a simplified (demonstration) example of information systems' interaction in Russia. A special feature of this example is the use of really acting services that open the access to the information systems of the state agencies.

In developing this example, we used the following scenario: the X-ray specialist applied to the branch of the Federal Social Insurance Fund (FSIF) territorial body (hereinafter referred to as the Branch) assigning for monthly payments in connection with an occupational disease. To make a decision on the payment, it is necessary to check the information containing in a number of documents. Therefore, before making a decision, the Branch requests the confirmation of some information from the other agencies. The Ministry of Health (Source 1) is a supplier of the

[1] The program Digital Economy of the Russian Federation approved by the Government of the Russian Federation in its resolution No. 1632-r, July 28, 2017, Available at: http://ac.gov.ru/files/content/14091/1632-r-pdf.pdf.

"information about the medical worker and his professional work activity"[2]. The Federal Medical Biological Agency (FMBA, Source 2) provides the "information on the existence of a causal relationship between developed diseases and the disability caused by the effects of radiation exposure". We request this data set from the Cross-agency Electronic Interaction System of St. Petersburg[3]. In accordance with the requested data, the information is checked.

In the example, we show the solution for the task of integrating the service of the Federal Register of Medical Workers (Source 1, MedStaff Web Service) and FMBA service (Source 2) with the branch information system. We assume that the Branch has an information system that performs all the necessary business processes of its activity (including checking information in the documents provided by the Applicant).

In the classical version of integration – the architecture with a mediated schema [6] – (Option 1), in the information system of the Branch we implement a unique client of each web service (hereinafter referred to as the client) to request and obtain the necessary data (Fig. 1). The client contains the program code for calling the corresponding web service (WS), requesting and converting the data from/to the data of the Branch. To simplify the schemas, the Branch's clients, as well as MedStaff WS and FMBA WS, encapsulate all software layers, including references to the relational DBMS. In this case, information systems have their own independent (and often contradictory) data models, which become the basis for the formation of a mediated data storage schema.

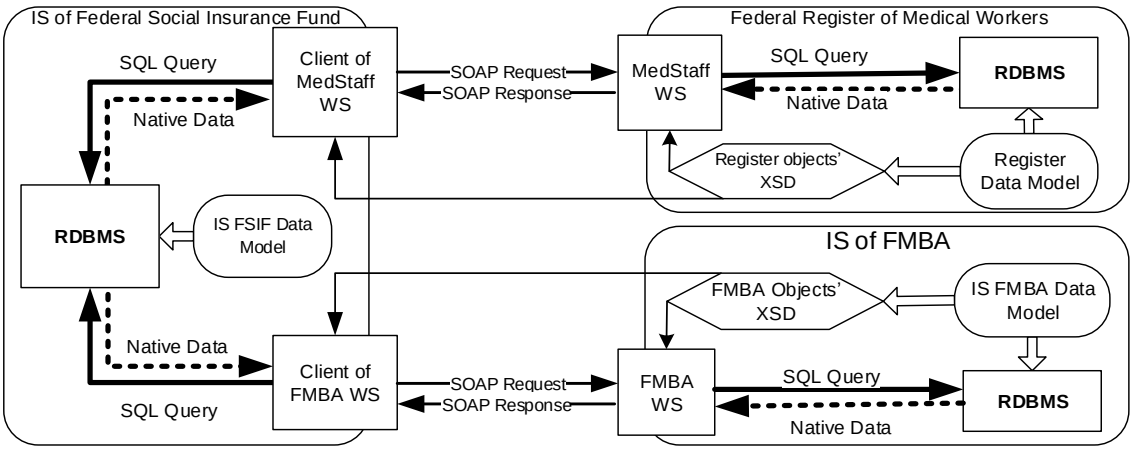

Figure 1. Classical integration of data

Among the obvious disadvantages of this integration method, we point individual development of the client for each web service, as well as the need to re-program the client code every time, when the changes in web services occur due to the changes in the data model or a storage schema.

Modern approaches to the integration of heterogeneous sources based on the construction of an integrated data model [6] (Option 2), for example, NIEM [5], provide independent development and maintenance of a unified data model (UMD) (Fig. 2). For the information exchange, it is enough to "know" two models: UMD and the own model of the information system. UMD is a common/reference model for the information sharing and exchange, but the data models of information systems are conducted locally (Fig. 2). Thus, the IS data model can be created both with and without UMD. This is especially important for the integration of legacy information systems and helps to avoid a significant and costly upgrade.

[2] http://ms-info.rosminzdrav.ru/index.php?option=com_content&view=category&layout=blog&id=69&Itemid=101
WSDL: http://ms-info.rosminzdrav.ru/attachments/article/323/MedStaffWCF_WSDL_XSD.zip
[3] https://smev.spb.ru/SMEVRegistry2/viewService.jsp?serviceId=233. WS passport: https://smev.spb.ru/SMEVRegistry2/UploadServiceFile?fileId=10034&type=2

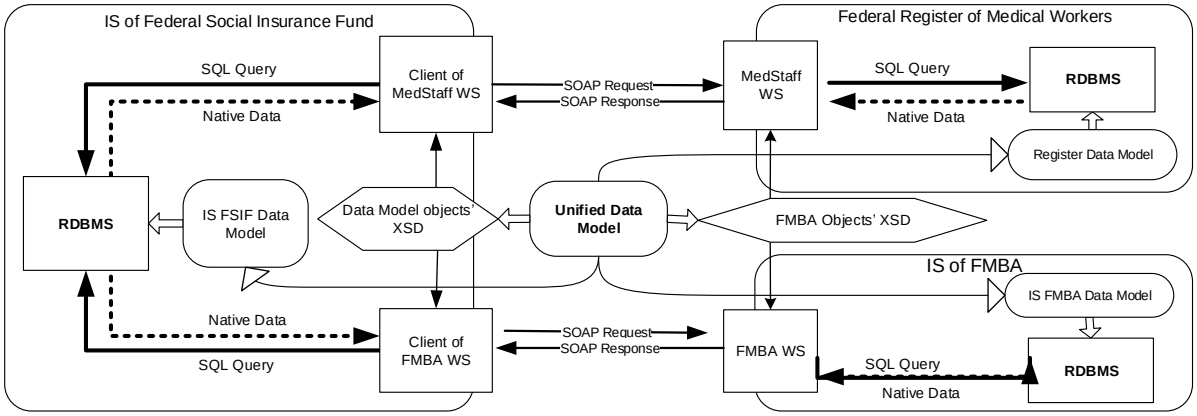

Figure 2. Local use of a unified data model in the processing system

The option with UMD-based integration provides an unambiguous interpretation of domain objects by all interaction participants and allows eliminating some drawbacks of the classical method (Option 1). When developing the client and the web service itself we provide the compliance to UMD but not to the data models of multiple systems. It is clear, the UMD change causes the changes in the client code as well as in the web services, but changing the IS data models of the Sources while keeping the UMD unchanged entails the modification only of the web services, but not the clients.

Applying semantic integration [6] (Option 3), we create an ontology-based model (domain model) which describes the subject area. It is also a common/reference model (Fig. 3), but unlike option 2, all the participants of the information exchange use it without the possibility to make local changes in the IS data models (excluding internal extensions).

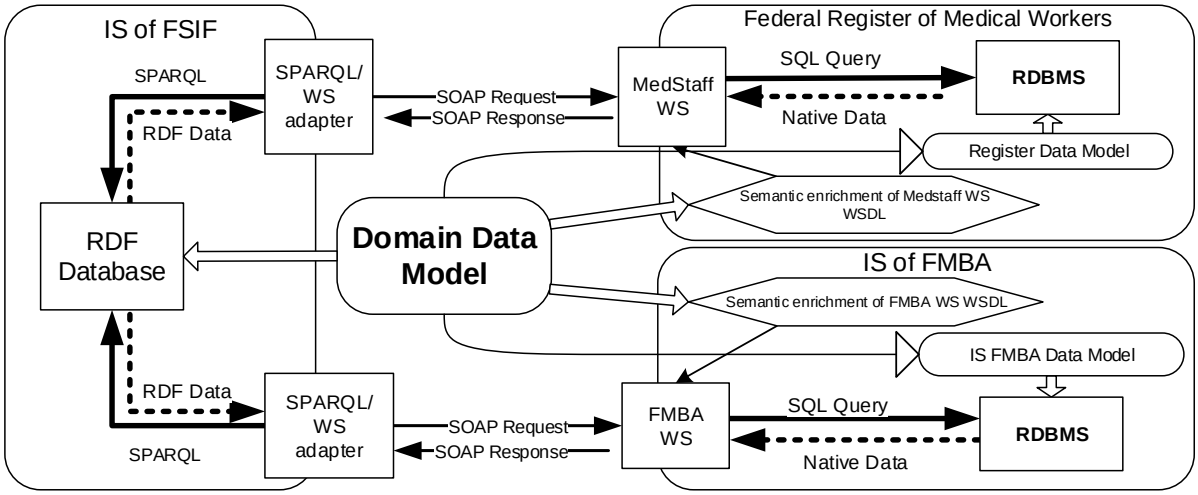

Figure 3. Use of domain model by all the information exchange participants

The advantage of this approach is that UMD changes do not cause multiple modifications in the mediators (Option 1) or web service clients (Option 2), because the configuration of SPARQL adapters goes automatically in compliance with the domain model. In addition, the application of the domain model allows the use of semantically oriented tools (for example, ontologies) to control the integrity and consistency of the interaction.

Digital Health ecosystem is an open system of systems [7] with a changing number of interaction participants [8], which entails frequent changes in the data model. Therefore, the application of the first method becomes extremely inefficient, and the use of UMD will lead to constant improvements in the clients of web services. Thus, the advantages of semantic integration for building the Digital Health ecosystem are very distinct.

3. Example of applying semantic integration methods

In practice, we confirmed the advantages discussed above by implementing the example of semantic integration in healthcare within the project of Plekhanov Russian University of Economics named the "Center of semantic integration"[4].

At the stage of preparation, we uploaded various types of semantic assets (SA) to the catalogue of SA, based on ADMS [12]. We described the data models of web services (input and output data schema), which then we used in the developed example. We registered the models in the SA catalogue and fixed their association with the information about the services (links to web sites, passports, etc.). Each element of WS data model contains a description of the element, its type, description and other attributes, including the link to relevant vocabularies registered in the SA catalogue.

The developed domain model contains the following objects: "Person", "Disease", "Medical institution" with the required attributes (Fig. 4). We linked created elements with the corresponding elements of other SAs, for example, vCard, FOAF and other.

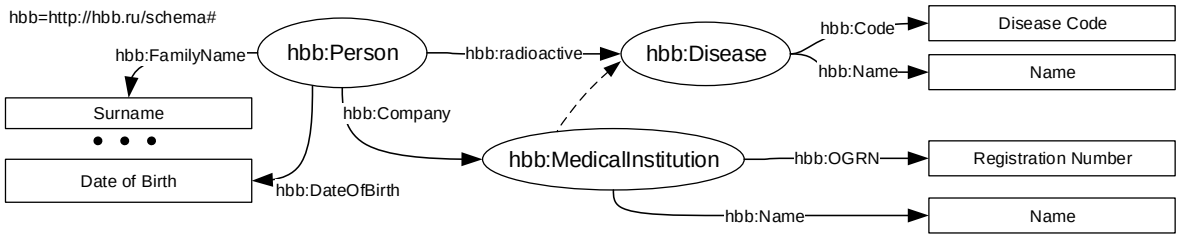

Figure 4. Example of the domain data model

To meet the features of the implementation, we extended the basic domain model in correspondence with WS data models by means of mapping and linking. As a result, the elements of the domain model are associated with the data model elements of the web services chosen for the example (Sources 1, 2) and other relevant schemas. Association of WS data model elements with the elements of the domain model provides the same interpretation of data during the information exchange[5].

Application of the semantic approach to the organization of interaction and data storage in the form of RDF objects (using the model "object" – "attribute" – "value") provides the possibility of sharing the information from different sources. At the same time, it helps to enrich the information exchange with the additional opportunities for the data search using the links that are not explicitly specified. For example, the classical integration method makes impossible to find the information about medical facilities in which workers received a disease or disability from radiation (in Fig. 4 the broken line indicates this link) without downloading all the data, because this link is not available in the web service. Whereas a simple SPARQL query will allow to obtain this information without a revision of the domain model.

The enrichment of WSDL schema with semantics can transfer WS to the category of a semantic web service. This provides the possibility of inheriting the already existing service infrastructure. We can also combine a set of semantic web services into a semantic integration bus, in which the central part is the domain model, and the services themselves contain all the necessary information for semantic integration and interpretation of data.

4. Conclusion

The semantic approach drastically reduces the need for constant updating of cross-system interaction interfaces and its implementation for the information sharing in a variable heterogeneous

[4] Centre of Semantic Integration, Portal: http://csi.semanticpro.org/catalog
[5] Additional materials for the example: http://csi.semanticpro.org/library/article/172.ru.html

environment is important. Despite an open number of participants, changes in technologies and software platforms, as well as in the conditions of information exchange, it provides the access to the systems and services within the open connection of independent consumers. Semantic integration extends the possibilities of obtaining information about the relationships between the elements that are not explicitly specified.

The use of semantic approach to the integration provides significant advantages and allows the formation of a semantic core, which will keep the knowledge and become the basis for the Digital Health ecosystem [8].

References

[1] Serebryakov V.A., Semantic integration of data, presentation 2012, http://sp.cmc.msu.ru/proseminar/2012/serebryakov.2012.04.20.pdf

[2] Chernyak L., Data Integration: Semantics and Syntax, Open Systems №10, 2009, http://www.osp.ru/os/2009/10/11170978/

[3] C. J. Date, An Introduction to Database Systems (8th Edition). Pearson Education Inc., 2004, p. 1024, ISBN 0-321-18956-6

[4] Madnick S., Gannon T., Zhu, H., Siegel M., Moulton A., Sabbouh M., Framework for the Analysis of the Adaptability, Extensibility, and Scalability of Semantic Information Integration and the Context Mediation Approach, Massachusetts Institute of Technology Cambridge, MA, USA, 2009

[5] Akatkin Yu.M., Yasinovskaya E.D., Shilin A.V., Bich M.G., Methods of Semantic Integration in Distributed Information Systems: Challenges of Application, Selected Papers of the 7th International Conference Distributed Computing and Grid-technologies in Science and Education Dubna, Russia, July 4-9, 2016.

[6] M.R. Kogalovsky, Methods of data integration in information systems, Moscow 2010, http://www.ipr-ras.ru/articles/kogalov10-05.pdf

[7] Karpov O., Akatkin Yu., Konyavsky, V. Mikerin D. Digital Health in Digital Society. M.: Delovoy Express., 2016. - 492 c. ISBN 978-5-89644-126-7

[8] Karpov O., Akatkin Yu., Konyavsky, V., Shishkanov D., Yasinovskaya E., Digital Health in Digital Society. Ecosystem and Cluster, Moscow 2017.

[9] NIEM USER GUIDE VOLUME 1 May 20, 2008 URI: http://reference.niem.gov/niem/guidance/user-guide/vol1/

[10] Yu. Akatkin, E. Yasinovskaya, V. Drozhzhinov, What is NIEM // http://www.cnews.ru/reviews/new/ikt_v_gossektore_2014/articles/chto_takoe_sistema_mezhvedomstvennogo_vzaimodejstviya_niem/

[11] V. Peristeras, N. Loutas, S. K. Goudos and K. Tarabanis, "A conceptual analysis of semantic conflicts in pan-European e-government services,"Journal of Information Science, vol. 34, pp. 877-891, 2008.

[12] Akatkin Yu.M., Yasinovskaya E.D., Shilin A.V., Bich M.G. Management and (re)use of semantic assets for information sharing – M.: 2016. 428 p., pp.235-242 ISBN 978-5-94588-206-5, http://elibrary.ru/item.asp?id=27401010