SEMANTIC INFORMATION MANAGEMENT: THE APPROACH TO SEMANTIC ASSETS DEVELOPMENT LIFECYCLE

Yu. Akatkin [a], E. Yasinovskaya [b], M. Bich [c]

Plekhanov Russian University of Economics, 36, Stremyanny lane Moscow, Russia 117997

E-mail: [a] uakatkin@semanticpro.org, [b] elena@semanticpro.org, [c] misha@semanticpro.org

The application of semantic integration methods faces challenges arising at collaboration between IT-specialists and domain experts during the model building stage. These challenges can affect the correctness of domain formalization as well as the whole outcome of the integration in distributed information systems. To overcome the lack of semantic interoperability we suggest the creation of a collaborative platform which provides the (re)use of semantic assets (SA) for further semantic integration. The analysis of the limitations existing in SA management standards leads the authors to propose the collaborative approach, based on an extended lifecycle of semantic assets. The authors consider the implementation of the platform based on the Asset Description Metadata schema extension to be a rational option.

Keywords: Semantic interoperability, semantic information management, lifecycle of semantic assets, semantic integration, distributed information systems, ADMS, model driven architecture

1. Introduction

Lots of scientific papers describe the application of semantic methods in the integration of data from heterogeneous sources [1, 2, 4, 9, 10, 12, 14, 17]. Currently distributed systems mostly support interoperability on technical and organizational levels as a rule. Although, semantic (information) interoperability becomes essential for successful integration. The effective implementation of semantic methods for the development of e-Government in Europe and the United States[1] over the last 15 years [6, 12, 14] has proved this statement in practice.

The ability to understand the meaning of data determined by the provider becomes extremely important. It improves the growth of semantic interoperability significance in heterogeneous environment with constantly changing number of participants. Data-centric paradigm is now the key enabler for the development of disruptive technologies in a new digital world [7, 16].

The use of semantic data models [6] (e.g. glossaries, dictionaries, taxonomies, thesauri and ontologies) – hereinafter referred to as semantic assets (SA) [13] – is the basis for semantic integration. SA enable semantic interoperability of distributed information systems (IS) and serve for data collection, search, analysis and data visualization performed by using semantic properties.

As a rule, groups of domain experts construct SA throughout the process of information system development or modeling of a particular domain. It is important to transfer the domain knowledge from paper documents into machine-readable formats (ontologies, thesauri, and glossaries) and to provide its dissemination outside of a specific information system. It brings the "understanding" of data to the information exchange and simplifies further SA (re)use by other IS interacting in a heterogeneous environment. At the same time, already developed SA are often insufficient for the modeling of information exchange, therefore, experts come across the task to build a more detailed semantic exchange model based on SA applied to the domain of interacting IS.

Existing ontology-based approaches for semantic interoperability have not been sufficiently effective because "there is no systematic methodology to follow, no concert methodology for building ontologies and all existing ontology-based not able to reconcile all types of semantic conflicts". Ontology Summit supports this experts' concern and emphasizes that "in practice, however, Semantic Interoperability is difficult to achieve" [11].

We consider it necessary to join the efforts of IT-specialists and various domain experts providing the ability of information systems to interact on a semantic level. Additionally, their cooperation should help to solve the problem of cross disciplinary misunderstanding which results in multiple revisions and unsatisfactory results. To overcome the challenges shown above and to simplify the application of semantic integration methods at the stage of semantic asset development we suggest setting an expert-oriented, common methodology for SA Management and support it with appropriate tools.

2. Semantic Assets Management

Semantic Web expansion have caused the development of many semantic assets that have become standards and recommendations, which de facto or de jure define various components or data schemes (e.g. Dublin Core, FOAF, VOAF, SKOS, vCard, etc.). The need to use solutions supporting accessibility and (re)use of semantic assets, considering special aspects of localization, has become obvious for the accumulation and spread of knowledge encapsulated in SA. Within the general guidelines some projects (e.g. JOINUP [12, 13]) stipulate the methodology for semantic assets' management. However, existing SA cataloguing platforms do not provide a complete lifecycle especially at the stages of SA development and modernization.

Within the framework of the European Interoperability program ISA [6] in 2011-2012 the ADMS Working Group [12] developed the Asset Description Metadata Schema (ADMS)[2] to collect, search and study the compatibility of semantic assets. In 2013, the extended ADMS (ADMS-AP 1.0) specification was developed. This profile was focused on semantic interoperability based on the

[1] NIEM 2017. National Information Exchange Model, https://www.niem.gov
[2] W3C 2013. Asset Description Metadata Schema (ADMS). W3C. http://www.w3.org/TR/vocab-adms/

unification of SA descriptions. In 2016, following the use experience of ADMS-AP 1.0, the revision of the specification resulted in the adoption of the version of ADMS-AP 2.0 [5]. ADMS and, accordingly, JOINUP platform, describe SA's lifecycle using four statuses: Completed, UnderDevelopment, Deprecated, and Withdrawn. We consider the use of ADMS for SA cataloguing, as well as, for storing and publishing their descriptions. But we suppose in practice this lifecycle includes some additional steps and should be extended for the following stages of SA development: (1) experts' collaboration, (2) validation, (3) assessment and quality evaluation of represented semantic assets.

We have also studied W3C recommendations [15], and ISO/IEC 11179 [8] and the conclusion is both support the lifecycle but manage the entity as a single document. The adoption of these approaches is reasonable only for SA management without considering the constituent elements. However, at the stage of SA development/modernization the contents of semantic asset, i.e. its elements and their properties, play the main role.

Therefore, we propose to combine the lifecycle management of semantic models as a single asset (e.g. ADMS repository records) with the change management of their contents. The reason for this consolidation is the ability of SA contents to split into parts (branches, sections, sets and elements), each of which could follow its own specific workflow and should be controlled, reviewed and assessed during the stage of development.

We think it reasonable to use a widespread and recognized by web community W3C lifecycle to describe the development stages of SA (ADMS status "In Development"). It reflects the features of an open SA catalogue such as attracting a wide range of experts, a variety of SA domains, use of SA in Semantic web and exchange of information using web-based technologies.

3. Semantic Assets Development Lifecycle

To bridge the gap associated with the lack of understanding between domain and IT experts in the development of information systems and their interaction, a collaborative semantic integration platform [3]. Working on the SA lifecycle management methodology during our R&D initiative to create the Center for Semantic Integration (CSI) at the Plekhanov Russian University of Economics we revealed and covered the following scenarios:

- **Domain experts** develop semantic assets in scope and detail necessary for IT specialists during SA implementation in information systems.
- **Domain experts** establish the correspondence between different SA elements to harmonize their content.
- **IT specialists** ensure the consistency of the developed semantic models with the description of the subject area, generally accepted standards and recommendations regarding the composition and contents.
- **Domain experts** review the models developed by IT specialists to assess semantic completeness and consistency.

Basing on these scenarios we set the primary task to support both (1) the development of semantic assets and (2) the harmonization of SA contents. We register existing semantic assets in the catalogue, based on ADMS. After expert review they can be loaded from external resources for further (re)use (e.g. for localization). The level of expert review is selected depending on the level interest: (1) validation among a working group, (2) public expert assessment by the community, (3) an independent review conducted by domain experts.

Throughout the collaboration of working group participants during the SA development, various stages can occur (see Figure 1), which can be described by W3C workflow statuses.

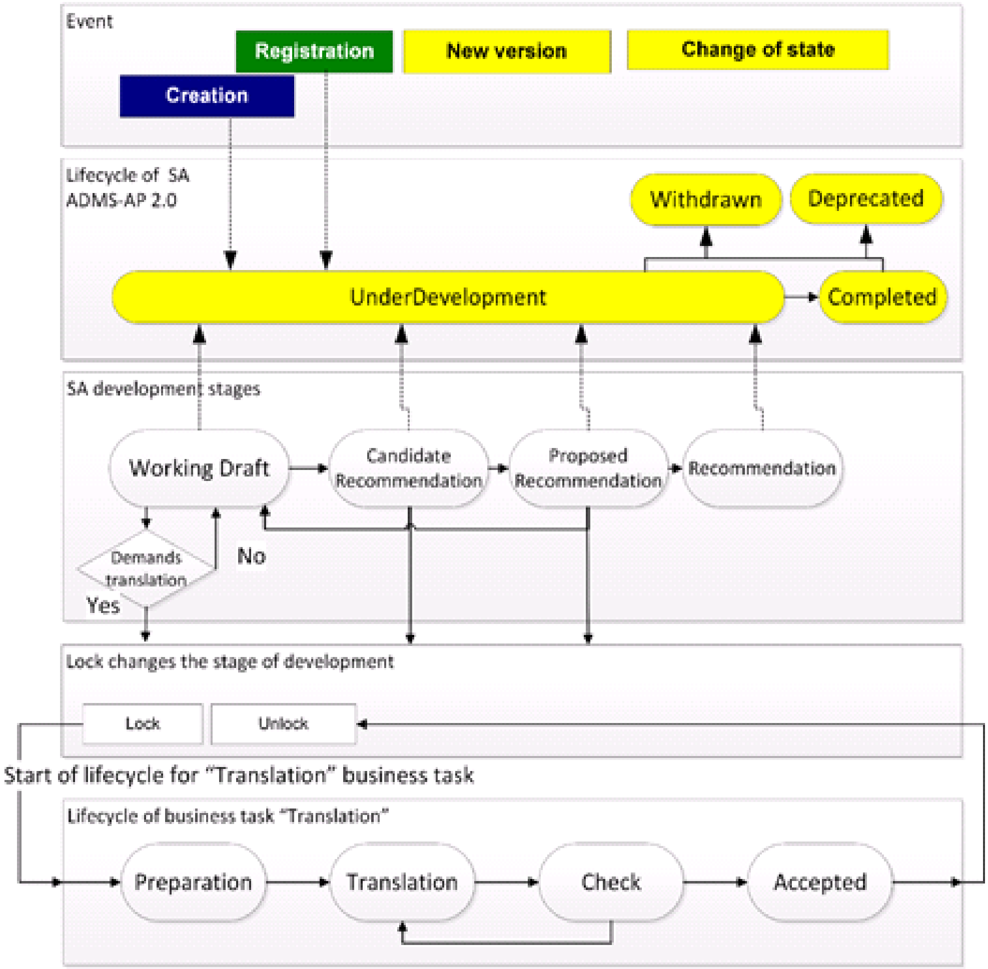

Figure 1. Lifecycle of semantic assets

We offer to arrange the process of SA development by assigning business tasks and controlling their performance. In Figure 1, the solid lines indicate the possible transitions between the lifecycle states or the development stages; and broken lines show the connections between the states, stages, and events that occur during the SA lifecycle. Each stage of SA development has a set of business tasks available for assignment (see Table 1).

Table 1. Stages of SA development stages and business tasks

SA development stages	Business tasks for SA development stages	Comments
Working Draft (WD)	SA Loading SA Creation SA Translation SA expert assessment	Filling the main content of the asset (loading SA contents, creation of new SA elements, translation, etc.) and refining the description of SA (completion of fields, classification, connection to other assets).
Candidate for recommendation (CR)	SA Modification SA expert assessment	Main expert evaluation and making small changes, e.g. spelling improvement.
Proposed recommendation (PR)	SA expert assessment	Decision that the SA can be recommended for implementation.
Recommendation (REC)		

Implementation of methods and tools for the reuse of semantic assets including constituent elements imposes restrictions on SA deletion/dismissal. In addition to the usual notification, informing SA users about the changes, the methodology should cover the events associated with the SA withdrawal (ADMS-AP 2.0 lifecycle status “Deprecated” or “Withdrawn”). The existing catalogues serve only for SA collection and provide several versions of a semantic asset existing in parallel. However, it is necessary to preserve the knowledge about the changes in semantics and the link between “old” and “new” semantic assets should be established.

To support the lifecycle of semantic assets we have developed the extension to ADMS-AP 2.0, named ADMS-WF 1.0. It provides compatibility and reuse of SA by domain experts and IT specialists at the stage of development. For the implementation of the usage scenarios represented above we offer to extend ADMS-AP 2.0 classes and add:

- additional properties to ADMS “Asset”, dcat:Dataset class for the conduct of SA versioning, access to SA contents together with the connection to business tasks and events.
- additional property to ADMS “Asset Distribution”, dcat: Distribution class for the integration with external design tools.

Besides that, it is necessary to add:

- “Event”, dct:Event class for the storage and distribution of events, occurring during SA development.
- “Change Request”, cm:ChangeRequest class in order to fix business tasks connected with SA as well as their workflow.

Special semantic models such as controlled vocabularies of events, business tasks types, design tools and SA lifecycle states determine available property values. ADMS-WF 1.0 supports backward compatibility of SA catalogues, implemented in CSI collaboration platform, with repositories, correspondent to ADMS-AP 2.0.

4. Conclusion

This article presents the approaches to the achievement of semantic interoperability in heterogeneous information systems environment. To provide an unambiguous, meaningful interpretation of data by all the participants of information sharing we keep to the following principles: (1) consolidation and reuse of semantic assets, and (2) the collaboration of domain experts and IT specialists working with semantic assets.

We consider the methodology for such collaboration mainly based on the reuse existing methods and Web standards. Due to the analysis of limitations inherent to known SA management standards we reveal the necessity to extend SA lifecycle and to support the expert workflow at the stage of SA development. Following the usage scenarios determined during our R&D at the CSI project we suggest combining lifecycle of SA (including the elements), SA development stages and lifecycle of business tasks.

To implement the proposed approaches, we have developed an extension to ADMS-AP 2.0 – ADMS-WF providing backward compatibility of the collaborative semantic integration platform with ADMS-compliant catalogues together. It also serves for information sharing and dissemination of SA built or modified in the process of experts’ collaboration for the semantic integration. Companies owning semantic assets, experts and IT specialists can use the ADMS-WF 1.0 profile for cataloging, as well as for supporting the lifecycle of semantic assets during harmonization, developing new or finalizing existing versions.

In addition, experts can use the ADMS-WF 1.0 to support the lifecycle of SA during the process of semantic assets discussion and assessment. IT specialists can use this profile to support the lifecycle of SA when validating the correspondence of data schemas (metadata) to domain models or enriching data schemas with the semantic information needed to support interoperability of heterogeneous information systems.

References

[1] Akatkin Yu.M., Yasinovskaya E.D., Bich M.G. 2016. Management and (re)use of semantic assets for information sharing. DAMDID / RCDL'2016, 235–242, http://elibrary.ru/item.asp?id=27401010

[2] Akatkin Yu.M., Yasinovskaya E.D, Bich M.G., Shilin A.V. 2017. Application of semantic integration methods on the example of web services information exchange arrangement. Information Resources of Russia №2 (156) – 2017, P.13-19

[3] Akatkin Yu.M., E.D. Yasinovskaya, Bich M.G 2016. Methods of Semantic Integration in Distributed Information Systems: Challenges of Application. 1787 (2016), 80–84. http://ceur-ws.org/Vol-1787

[4] Date C. J. 2004. An Introduction to Database Systems (8 ed.). Pearson Education Inc., Pearson Education Inc. P. 1024. ISBN 0-321-18956-6. P247

[5] European Commission 2016. ADMS Application Profile for solutions on Joinup Version 2.01. European Commission. https://joinup.ec.europa.eu/node/150484/

[6] European Commission. 2010. Annex 2 to the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions «Towards interoperability for European public services», COM (2010) 744 final.

[7] Fernández, A. 2018. Putting value at the core of your business: Data-driven vs Data-centric, Stratio, http://www.stratio.com/blog/datadriven-versus-datacentric/

[8] International Organization for Standardization. International Standard ISO/IEC 11179-6:2015, Information technology - Metadata registries (MDR) - Part 6: Registration. International Organization for Standardization

[9] Madnick, S., Gannon, T., Zhu, H., Siegel, M., Moulton, A., Sabbouh, M. 2009. 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 http://web.mit.edu/smadnick/www/wp/2009-02.pdf

[10] Madnick, S.E., & Zhu, H. 2006. Improving data quality through effective use of data semantics, Data and Knowledge Engineering, 59(2). P.460-475 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.3747&rep=rep1&type=pdf

[11] Ontology Summit. Ontologies within Semantic Interoperability Ecosystems. In Ontology Summit 2016 Communique. http://ontologforum.org/index.php/ OntologySummit2016/Communique

[12] Vassilios Peristeras, Gofran Shukair, Nikolaos Loutas. 2012. Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema. Computers in Industry 64 (2013) (2012), 10–18. https://www.sciencedirect.com/science/article/pii/S0166361512001406

[13] Vassilios Peristeras. 2011. Federation of Semantic Assets Repositories. https://joinup.ec.europa.eu/sites/default/files/08/5e/bb/federation-of-semantic-assets-repositories.pdf

[14] Vassilios Peristeras. 2013. Semantic Standards: Preventing Waste in the Information Industry. IEEE Intelligent Systems, 72–75. https://doi.org/10.1109/MIS.2013.115

[15] W3C. W3C Consortium Process Document. W3C. http://www.w3.org/Consortium/Process/

[16] Wahlin, R.: Data-centric architectural best practices: Using DDS to integrate real-world distributed systems, Military Embedded Systems, http://mil-embedded.com/articles/data-centric-real-world-distributed-systems/

[17] Walaa, S. Ismail, Mona M. Nasr, Torky I. Sultan, Ayman E. Khedr. 2013. Semantic Conflicts Reconciliation as a Viable Solution for Semantic Heterogeneity Problems. (IJACSA) International Journal of Advanced Computer Science and Applications. Vol. 4. No.4 https://pdfs.semanticscholar.org/22bb/9c1c27049fce16d1546c54d51e0ede2854d8.pdf