Towards inter-operable enterprise systems – graph-based validation of a context-driven approach for message profiling

Elena Jelisic[1], Nenad Ivezic[2], Boonserm Kulvatunyou[2], Scott Nieman[3], Hakju Oh[2], Sladjan Babarogic[1], Zoran Marjanovic[1]

[1]Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia
{elena.jelisic,sladjan.babarogic,zoran.marjanovic}@fon.bg.ac.rs
[2]National Institute of Standards and Technology, Gaithersburg, MD, USA
{nivezic,serm,hakju.oh}@nist.gov
[3] Land O'Lakes, Shoreview, MN, USA
stnieman@landolakes.com

Abstract. Providing the business context has a potential to become a powerful mechanism for the interoperable usage and efficient maintenance of message standards. In the literature, there are multiple techniques for its representation and application. Industry use cases have identified multiple issues that come with currently used techniques, which can represent that context. Initial assessment of a logic-based technique for doing so has been conducted and showed that some of the identified issues can be resolved. This paper presents plan for validation of the logic-based technique where proposed algorithms will be assessed in realistic integration scenarios. The paper gives details of the validation steps, goals that need to be met, and indicates issues that guide our future research plans.

Keywords: Business context, Services integration, Enterprise interoperability.

1 Introduction

Distributed and autonomous services are increasingly becoming components of the manufacturing and logistics enterprises. To achieve efficient integration of these services, message standards are needed. Messages (i.e., payload or business document specifications) define types of transaction-related information that need to be exchanged among different services. Messages are necessarily standardized because of the wide variety of inputs to those services. Inputs come from numerous business sectors, business processes, business contexts, and business representatives.

However, traditional message standards have been only partially successful. We have identified four major issues inhibiting the adoption of messaging standards. First, "out-of-the-box" message standards are necessarily generic, making them difficult to adapt to specific use cases. Second, the standards are typically documented in text form to facilitate human interpretation and do not have a computational form that can be processed by validation tools to assure integration and interoperability. Third, over time the standards become large supersets of duplicative data elements contributed by various industries making the standards difficult to implement. And, fourth, as more elements are added, detailed refinement or profiling of a message standard is necessary to recapture the original business intent and business context.

On-going work by the Open Application Group [1], industry, and National Institute of Standards and Technology (NIST) [2] continues to explore the concept of using business context to improve 1) the enablement of new syntax-neutral message standards and 2) the resolution of the identified issues [3, 4, 5]. Some of the currently implemented context schemas and management approaches, however, have proved to be inadequate [6]. On the other hand, some newly proposed, promising approaches are introduced [7] but must be tested with realistic industry situations to establish their capability and feasibility. Yet, such testing is resource-intensive and requires commitment and engagement of industry and standardization communities.

We have started to validate a newly proposed approach using existing scenarios already contributed to OAG-i by industry stakeholders [1, 6]. Further industry adoption of the proposed context management approach depends on its capability to address all real-world use cases. Accordingly, the theme for this paper is our proposed approach to validate a promising, context-management technique for cross-industry use cases. Since the outcome of our approach depends on the validation process, this paper provides a detailed description of that process. That description identifies the goals and steps needed to achieve those goals. To the best of our knowledge, there are no research papers that deal with validation of a proposed, business-context-management approach.

In the following, Section 2 introduces background information about important concepts that are used in the paper, Section 3 describes the validation process through steps needed to achieve identified validation goals, Section 4 describes validation domain, and, finally, and Section 5 discusses the expected results and future research steps.

2 Background

Message standards are expressed in terms of message components. Message components are developed and offered as parts of a message standard suite where they are used to form multiple types of messages. An important, international standard, Core Component Technical Specification (CCTS), has been proposed to enable uniform and consistent development of message components and message standards. The goal of CCTS is to advance interoperability among applications and services [8].

A key concept in CCTS is *business context*, which is used to describe the business intent of using a message component or message profile, which is a subset of a corresponding message standard. In our view, capturing that intent is essential to finding and reusing the right message components and profiles, which will enable interoperability of corresponding data and services. Following CCTS, a specific business context is represented as a set of context values that are associated with their corresponding context categories (e.g., business process party profiles, geo-political location, tasks with a business process, resources available, and industry).

While the business context can indeed identify situations in which some components will be reused, we found this to be insufficient for interoperability. This is because semantic interoperability must also consider what happens in the background (e.g., transformation of the data into the database). Hence, in this paper, we assume that business context must also provide semantic definitions and validation rules that may be triggered when choosing the components in the business context. Providing those definitions and rules requires business-context knowledge.

Business-context knowledge can be represented using multiple, modeling techniques including graphical, object-oriented, logic-based, and ontology-based. For the purpose of this paper, we will employ a logic-based modeling technique - Enhanced UN/CEFACT's Context Model (E-UCM) [7]. E-UCM represents business-context knowledge using decentralized, directed, acyclic graphs. Each business-context category has at least one such graph that represents a list of possible values (or nodes) in the corresponding category. Then, a specific business context for a specific, information-exchange situation corresponds to a collection of nodes belonging to specific categories in the E-UCM graph.

E-UCM methodology for creating these graphs has two main parts. First, there is an infrastructure for (a) business-context presentation and expression and (b) contextualization of existing message profiles. Contextualization is supported by algorithms that can detect message components that are either relevant or irrelevant for a *Requested business context*. Second, there is a set of algorithms that can be used to generate the structure of new message profiles using existing message components that are relevant for *Requested business context* of a new message profile.

To determine a business context for a specific, information-exchange situation, we identify a collection of graph nodes that make up that business context. A list of specific nodes can be identified from E-UCM graph using E-UCM-provided expressions that are built using available operators and predicates. Every message component and profile has an associated E-UCM expression that captures its usage intent. This expression is called *Assigned business context*.

Along with the *Assigned business context*, E-UCM introduces terms *Overall* and *Effective business context* and provides directives for calculation of these terms for all situations and components, ultimately resulting in *Effective business context* for each component. The *Overall business context* can be understood as cumulative business context of some compound message component, while the *Effective business context* should reveal message component's relevancy in a *Requested business context*.

The process of assigning contexts to a message profile and its components, and calculation of *Overall* and *Effective business contexts* is called *contextualization*.

3 E-UCM Approach – On-going Validation

Our initial validation [6] concluded that E-UCM methodology provides a powerful mechanism for business context presentation and expression. Our goal in this paper is to validate the algorithms that E-UCM proposed for both *contextualization* of existing message profiles and profile generation. Our approach to achieve that goal involves two steps. The first step is to assesses *Effective business context* as sufficient evidence of message component's relevancy or irrelevancy in the defined business context. The second step assesses the algorithms for message profile generation as capable of constructing a valid, message profile for a *Requested business context*. The key difference between those two goals is whether message profile structure is available or not. If it is available, profile refinement occurs for new integration scenarios. Otherwise, message profile structure should be generated using E-UCM algorithms on the basis of existing message profiles.

We made two important assumptions in completing the validation study. First, we assumed *contextualization* would be realized by a specialized service provider. Such a service provider would need domain knowledge needed for the development (or selection from an existing standard library) and use of message components. Such domain knowledge would typically be obtained through 1) business analysis of existing business processes and 2) information exchanges that rely on documentary standards and standard operating procedures.

Second, we assumed that there is one contextualized *Initial profile* upon which all others are based. To obtain the *Initial profile*, a service provider would construct a special kind of domain knowledge that contains an *initial collection of document components* and a *set of business rules* that govern the usage of those components. Using this domain knowledge, the service provider would generate an output in the form of a contextualized *Initial profile* with its components assigned intended business context, corresponding to the business rules.

Our validation process is shown in Figure 1. First three steps are prerequisite and the same for both validation goals, fourth and fifth steps reveal our validation goals, while the sixth step is an assessment of validation results. The third step gathers validation cases that will be used to accomplish our, two, identified, validation goals. Each validation case is described using *Requested business context*.

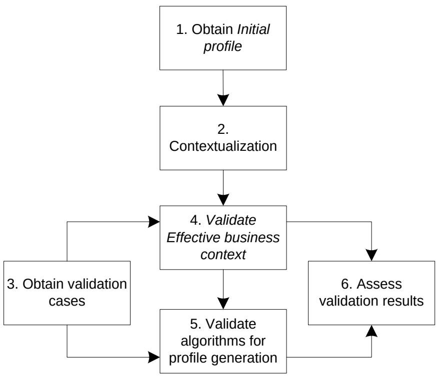

Figure 1 Validation process

3.1 Validation of *Effective business context*

The input for this validation step is semi-contextualized *Initial profile* with defined *Assigned business context* and calculated *Overall business* context and a set of validation cases. The output is a set of identified message profiles obtained through refinement of an *Initial profile.* The first step in our validation approach is to put the *Initial Profile* to use in a variety of validation cases, both boundary and normal, where business context is adequately and precisely expressed. For each validation case, *Effective business contexts* for each component will be calculated in order to identify irrelevant components. *Effective business context* is calculated as intersection between component's *Overall business context* and business context in which corresponding component is supposed to be used (*Requested business context*). Intersection is conducted for each business context category separately.

In general, there are three possible outcomes: *Effective business context* is equal to null, *Effective business context* is narrower than component's *Overall business context* and *Effective business context* is equal to component's *Overall business context*. *Effective business context* can be equal to null if intersection of the *Overall business context* and *Requested business context* expressions across any of their business context categories gives an empty set. This outcome informs us that the corresponding component or field is not relevant for the *Requested business context*. Other two outcomes inform us that the corresponding component is relevant, but it might undergo additional considerations (because *Effective business context* of the component is narrower).

The refinement of an *Initial profile* results in the identification of new message profiles for a specific set of integration scenarios, where only components that are relevant for the *Requested business contexts* are included. Message profiles' refinement results in recalculation of *Overall business contexts*. New message profiles will have *Assigned business context* equal to *Requested business context*. Figure 2 presents a reusable Contextualization and profile refinement subprocess that can be applied to any domain. Since the calculation of *Overall business context* depends on message component's type it is represented currently as a subprocess and details are neglected.

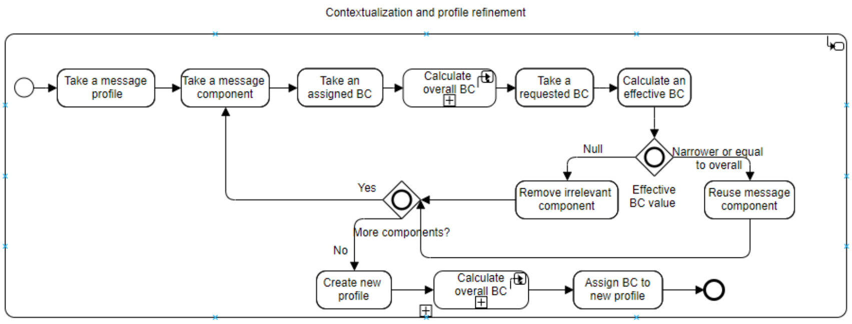

Figure 2 Contextualization and profile refinement process

3.2 Validation of algorithms for message profile generation

A set of refined message profiles, obtained from the previous step, and a set of validation cases provide the input for this validation step. An output is a set of generated new message profiles constructed using E-UCM proposed algorithms. In this validation step, we will take one by one validation case, and compare *Requested business context* with business contexts assigned to existing profiles that we got as the output from the previous validation step. If there is a match, we will reuse it, otherwise, we will employ E-UCM algorithms to obtain components, from the repository of contextualized ones, that are relevant for a new validation case. E-UCM proposes an algorithm with five processing steps. The first step gathers all components from an existing standard component library, while each of the next steps has the goal to filter message components applicable to the *Requested business context* regarding different matching indexes, such as complete or partial matching. Details of this algorithm are out of scope of this paper and will be discussed in our future work.

3.3 Results assessment

To achieve the identified goals, the structure of resulting profiles, both from the first and the second step, will be compared with expected outcomes by consulting human expert or developer. Important remark is that the fifth validation step is conducted under assumption that the output of the fourth step is valid.

4 Validation domain

We demonstrate our validation process using a case study from a Visa application domain. Afterwards, all conclusions will be verified in other domains, such as a Small-and-Medium-Enterprise (SME) and its procure-to-pay scenarios. We have identified three variables in Visa application domain: *Issuance country*, *Visa type* and *Applicant's country*. Moreover, we use these three variables as our business-context categories. Associated message profiles can include different combinations of these three. To simplify the complexity, we have fixed the *Issuance country=New Zealand* and *Visa type =Visitor*, respectively.

Since each country has its own, specific visa policies, we used a two-step process to construct a set of validation cases. First, for *New Zealand*, for *Visitor visa type* we identified a list of eligible countries whose citizens can apply for a visa. Second, for the same issuance country, for the same visa type we have identified a list of visa waiver countries. After these two steps are conducted, we can identify possible visa application submissions.

In the real life, New Zealand has one Visa Application Form (VAF) for Visitor visa type and this VAF, with the same structure, is offered to all applicants, no matter their country. According to our validation methodology setup, this *initial collection of document components* will be analyzed by the service provider, along with the set of New Zealand's specific business rules. As the output, we will get the contextualized *VAF Initial profile*. By studying the structure of visa applications for different visa types and different countries, we have concluded that there is a great chance that irrelevant fields will be proposed to a certain applicant. Korean application form has the field *Kanji Name* that is applicable for a few applicants' countries only. New Zealand's application form has the field *SWITCH card* which is the United Kingdom-specific credit card type, to name a few. These examples clarify the purpose of our first goal - to approve *Effective business context* as enough to identify relevant fields from the *VAF Initial profile*.

The output of this validation step would be a set of *VAF profiles* that would contain only components relevant to the *Requested business contexts*. Consequently, each applicant would get a customized VAF that contains only relevant set of fields, which will ultimately lead to less error-prone application process.

When a new visa application is recorded, first, we will compare *Requested business context* with business context assigned to existing profiles using the E-UCM business-context, matching algorithms. If there is a match, we will reuse it; otherwise, we will employ the E-UCM algorithm to construct a new message profile. That construction is based on relevant, obtained components from the repository of contextualized ones. The structure of resulting *VAF profiles* will be compared with expected outcomes by consulting human expert or developer. These conclusions contribute to the identified validation goals.

5 Discussion and Future Work

So far, first three validation steps have been conducted. Analysis of *Initial profile* and validation cases show that, in most situations, *Effective business context* can be used to identify irrelevant components from existing message profiles. However, there are some boundary scenarios where *Effective business context* is insufficient to make conclusions. We have identified the following types of unsuccessful validation cases. They justify the purpose of our validation process. From those example scenarios, E-UCM validation must be conducted prior to its industry adoption.

Time-dependent - Values assigned to business-context categories denote the current state of an object. We identified scenarios where it was necessary to remember the history, not just the current state. For example, keeping history of applicant’s country transitions might be important to decide whether some field is applicable to the integration situation at hand. In this case, *Applicant's country* business-context category would be used to denote transition of the applicant's country, not just his current citizenship. Example transitions include changes in citizenship, holidays, temporary work, and student-exchange programs. In practice, this means that the result of *Effective business context* might vary through time. Hence, considering only one state, without considering history of change, may not be sufficient to communicate needed information for the relevant integration situation.

Unpredictable - There are scenarios that lack the logic needed to determine whether some field is relevant to the applicant or not. For example, field *Passport book number* can often be found in VAFs. The *Passport Book Number* may appear in a passport in addition to the *Passport Number*. Some countries may have this detail in some versions of their passports, while others may not. There is no traceable guideline that will inform us which passport versions contain this detail, so defining *Assigned business context* for such component would be a challenge. Further, it means that calculation of *Effective business context* might be impossible.

Insufficient - There are scenarios where business context does not give us enough information. For example, *Applicant's country* may not be enough because applicant's current residence or applicant's nationality, to name a few, may also affect appearance of some field. This means that the business-context categories that are chosen to describe business context are not enough; and, other relevant categories should be introduced. In practice, *Effective business context* can be calculated, but it can lead to conclusions that are inadequate for the specific integration situation.

If our validation process proves *Effective business context* as insufficient, which we believe is the case, future research will be needed. We plan to consider using new context categories and employing data mining techniques for business context definition to resolve those failing scenarios and enable adequate calculation of *Effective business context*. These approaches may help identify combinations of context categories expected to adequately define a specific business context [4]. Also, there is a possibility for extension of E-UCM methodology to capture history data through business context categories. A possible solution would be introducing state-machine diagrams for business context definition that would enable capturing multiple states (values for specific business context category) through time.

The described on-going validation covers algorithms that can be employed for profiles' generation when there is only one document type considered. Also, E-UCM offers other algorithms that can be applied for profiles' generation when there are two, or more, paired document types. Future research will consider validation of this aspect of E-UCM methodology as well.

Disclaimer

Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

References

1. The Open Applications Group Inc. https://oagi.org
2. The National Institute of Standards and Technology (NIST). https://www.nist.gov
3. Ivezic, N., Ljubicic, M., Jankovic, M., Kulvatunyou, B., Nieman, S., & Minakawa, G. (2017). Business process context for message standards. CEUR Workshop Proceedings, 1985, 100–111.
4. Jelisic E., Ivezic N., Kulvatunyou B., Anicic N., Marjanovic Z. (2019) A Business-Context-Based Approach for Message Standards Use - A Validation Study. In: Welzer T. et al. (eds) New Trends in Databases and Information Systems. ADBIS 2019. Communications in Computer and Information Science, vol 1064. Springer, Cham
5. Jelisic, E., Ivezic, N., Kulvatunyou, B., Jankovic, M., & Marjanovic, Z. (2019, September). A Two-Tiered Database Design Based on Core Components Methodology. In European Conference on Advances in Databases and Information Systems (pp. 350-361). Springer, Cham.
6. Elena Jelisic, Nenad Ivezic, Boonserm Kulvatunyou, Scott Nieman, Hakju Oh, Nenad Anicic, Zoran Marjanovic, Knowledge representation for hierarchical and interconnected business contexts, Enterprise Interoperability IX, Tarbes, France, 17. - 20. Nov, 2020
7. Novakovic, D. (n.d.). *Business Context Aware Core Components Modeling.* Retrieved September 28, 2018, from Publikationsdatenbank der Technischen Universität Wien: https://publik.tuwien.ac.at/
8. UN/CEFACT Core Components Technical Specification CCTS, version 3.0., https://www.unece.org. Last accessed 10 March 2020