On the use of ontology-based integration architecture in Cemaden's Natural Disaster Observational Network

André A. S. Ivo
System Engineering
CEMADEN - INPE
São José dos Campos, Brazil
andre.ivo@cemaden.gov.br

Sheila G. Ribeiro
System Engineering
INPE
São José dos Campos, Brazil
sheila.granato@gmail.com

Fatima Mattiello-Francisco
System Engineering
INPE
São José dos Campos, Brazil
fatima.mattiello@inpe.br

Rogério Ishibashi
System Engineering
CEMADEN
São José dos Campos, Brazil
rogerio.ishibashi@cemaden.gov.br

***Abstract*—This paper describes a study performed in the Brazilian governmental center, named Cemaden, responsible to monitor natural disasters and sending alerts to the vulnerable community involved. The target of our study is the complex architecture of Cemaden's Observational Network. It comprises constituent systems (CS) provided by other partners organisms, which operation and maintenance are out of Cemaden controlling. The correct functioning of those heterogeneous CSs represents a major challenge in this type of system. Thus the comprehension of dependability issues regarding the information flow in the integration of the CSs is essential. Our study compares the current integration architecture based on syntax rules with the use of an ontology-based integration. The advantages of the proposed architecture are discussed in a case study.**

***Index Terms*—dependability, ontology, natural disaster**

I. Introduction

Between 2007 and 2011, the natural disaster occurrences in Brazil were significatively increased, reaching an annual recurrence of catastrophic events never recorded before. In 2007, the life of approximately 2.7 million people were socially affecteds [1].

The interaction of two fundamental factors, **social** and **natural**, is essential to understand what Natural Disasters refer to. [2]: (a) the conditions of social vulnerability that result in the occupation, generally irregular, of the national territory; and (b) the occurrence of extreme natural phenomena (rain or absence of it) that can trigger an event (floods, landslides, droughts) with potential for damages [3]. Therefore, rains, floods, and landslides are not synonymous with disaster since different social groups are not exposed to the same dangers and do not have the same conditions to face and recover [2], [3]. In summary, a natural disaster is characterized by an extreme natural phenomenon in a region of social vulnerability.

In July 2011, with the edition of Brazilian Presidential Decree nº 7.513, the National Center for Monitoring and Early Warning of Natural Disasters (Cemaden) was born, with the mission of monitoring and issuing alerts of natural threats in mapped risk areas of Brazilian municipalities. Besides that, Cemaden also conducts research and technological innovations that could contribute to the improvement of its early warning systems. [4].

Cemaden operates 24 hours a day, without interruption, monitoring 958 municipalities classified as vulnerable to natural disasters (mainly floods and landslides) [5]. Among other competencies, sends natural disaster alerts to the National Center for Risks and Disasters Management (CENAD), assisting the National Civil Defense System that operates in the municipalities. Cemaden has an observational network composed of 5857 instruments for environmental monitoring (rainfall, humidity, moisture, temperature, river level, among others) installed in risk areas distributed throughout the country. Fig. 1 shows the Situational Room where the specialists makes the natural disasters monitoring.

Fig. 1. Cemaden's Situational Room

The observational network comprises a large operational dimensions, which involves a high number of Constituent Systems (CS), each with independent authority and leadership. It has geographic dispersion, monitors the environment through sensors, and influences the physical environment through alerts that could be integrated with civil defense systems (sirens, visual signals such as emergency lamps, activation of dam locks, redirection of water resources, etc.). Moreover Cemaden is assisted by human systems, either in the role of experts who provide feedback to the system, or in the role of considering information about vulnerable social groups.

Ensuring the correct functioning of heterogeneous CSs represents a major challenge in this type of system. The effect that eventual CS malfunction occurrence might provoke in the whole system needs to be known. Preventive maintenance mitigates risks. However, many CS maintenances are out of Cemaden controlling. Thus the comprehension of dependability issues regarding the information flow in the integration of the CSs is very important. In this perspective, this work presents a study that compares an integration architecture based on syntactic rules with an ontology-based integration architecture. The paper sections are organized as following: section II presents Cemaden concept of operation and the existing integration architecture; section III proposes the Ontology-based Integration architecture; section IV discusses the use of the proposed architecture in a case study; finally section V concludes the paper.

II. Cemaden's CONOPS

Cemaden's observational network is considered one of the largest national networks of Data Collection Platforms (DCPs), however it is still insufficient to monitor over 50 thousand risk areas estimated by Cemaden [5], that is, 1 equipment for every 10 areas. For better coverage of the national territory, Cemaden also uses observational networks from its partners, such as the National Institute for Space Research (INPE), Department of Air Space Control (DECEA), National Water Agency (ANA), Geological Survey of Brazil (CPRM), among others. Fig. 2 illustrates the operational concept of the Cemaden's Alert System, in accordance to the Federal Government's Natural Disaster alert protocol.

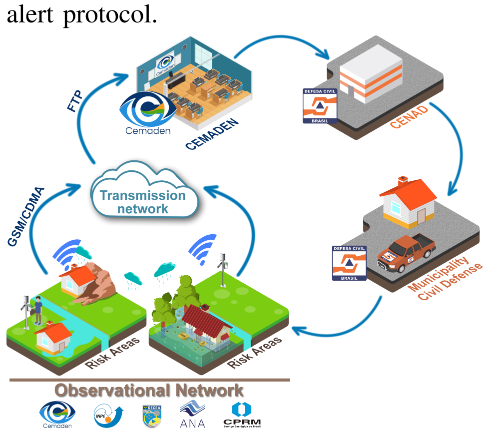

Fig. 2. Cemaden CONOPS - Concept of Operations

Cemaden's observational network is not a monolithic system or with fixed and well-defined dimensions, but rather a question of integrating heterogeneous systems in a coordinated manner to solve specific objectives, which makes the network a system with dynamic size and contours over time.

The partner systems integration (heterogeneous systems) is based on syntactic rules. Integration syntax refers to the rules set that regulate the information composition that will be exchanged between systems. It is observed that in this integration model, any change in syntactic rules can break the entire process, leading to failures. Fig. 3 presents the CEMADEN systems syntactic integration architecture.

In Cemaden's syntactic integration architecture, Fig. 3, syntactic rules are applied in the form of standardized text files. For these text files, rules are defined to identify the position of each relevant information to the integration process. The

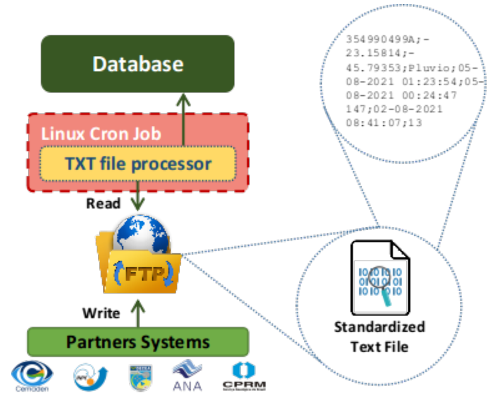

Fig. 3. Cemaden's Syntactic Integration Architecture

partner systems are responsible for writing the data files (on the Cemaden's FTP server) following the defined syntactic rules. Cemaden's processing systems read the files based on the same syntactic rules and then record the information in the database. This process makes evident the strong coupling and dependence between systems, significantly increasing their fragility, and lowering the dependability parameters.

III. Ontology-Based Integration

There are many means and techniques to accomplish this heterogeneous systems integration (Partner Systems or CS) the most popular and simple to apply are integrations based on syntactic approaches, such as the one currently used in Cemaden's observational network.

The composition of Cemaden's observational network is strongly based on heterogeneous systems with operational independence and not designed to operate together. It is in this context that semantic conflicts may occur. Semantic conflicts occur when the systems use different meanings for the same elements: they appear to have the same meaning but do not.

To reduce dependencies and coupling between CS, besides overcoming the semantic heterogeneities, the use of integration at the semantic level of information is proposed, as presented in the work of Izza (2005) [6], which proposes the use of ontology for the integration of systems such as ERP (Enterprise Resource Planning), CRM (Customer Relationship Management) and SCM (Supply Chain Management). Calhau (2010) [7] presented the Ontology-Based Approach for Semantic Integration (OBA-SI), which uses ontologies as conceptual models to define mappings between applications, and inspired this proposal for Ontology-Based Integration.

The concept of ontology is used in several areas of knowledge and with appropriate definitions for each one. The definition of ontology adopted in this work is as follows: *"An ontology is an explicit and formal specification of a shared understanding."* [8]. In this context, ontology refers to the meaning of an abstract model's elements, usually restricted to a specific domain. *Explicit* means that the categories of elements and their restrictions are explicitly defined. *Formal* refers to the fact that ontology can be read and interpreted by a machine.

The fact that an ontology is a formal theory about a given domain implies the need for a formal language for its definition, the so-called ontological languages [8]. There are several formal languages, such as SHOE (Simple HTML Ontology Extensions), DAML + OIL (Darpa Agent Markup Language + Ontology Interchange Language), RDF (Resource Description Framework), Web Ontology Language (OWL), which can be more or less appropriate to a particular domain of the system.

From the definition of requirements and the integration scenario, the specification process begins, in which semantic equivalences between applications are established. This step's main output is the specification of a system integration model, used to assign semantics to the systems' conceptual models and the integration process through mappings in the light of reference ontologies.

The use of ontologies throughout the development life-cycle allows the creation of shared knowledge from the early stages of the cycle, reducing the cognitive effort to understand the project among stakeholders. The Design and Development phases suggests using formal ontology languages to simplify the integration of heterogeneous constituent systems through an Integration Mediator.

A high-level semantic integration architecture implementation is shown in Fig. 4. This figure shows that the integration of a heterogeneous CS flows through the Integration Mediator component. The Integration Mediator's main objective is to translate the communication based on the semantic mapping presented by the Integration Model, an artifact produced in the analysis phase of the Life-cycle.

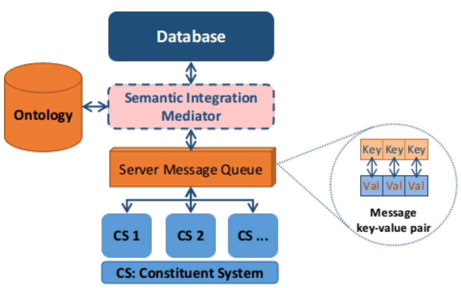

Fig. 4. Ontology-based Integration Architecture

In this work, the integration process is restricted to the application data (Payload) on layer 7 (Application) of the OSI Model (Open System Interconnection), where the main difference between syntactic and semantic integration architectures is coded.

In syntactic architectures, to obtain the meaning, it is necessary to know the syntactic rules used to interpret the data, that is, the data carries information, in addition, the business rules; in semantic architectures, the information meaning can be extracted from layer 7, which depending on the used protocol can separate the messages into key-value pair (KVP) and from the key obtain the semantic mapping, without efforts to interpret the data. A KVP is a set of two linked data items: a key, which is a unique identifier for some data item, and the value, which is the identified data.

The KVP model may not be available for some types of layer 7 protocols and, in this case, commercial and well-structured data models can be used, such as XML and JSON. Both XML and JSON are interpreted using syntactic rules, however, they do not carry the business rules, they are generic and suitable for any application.

As the business rules are not incorporated in the data, the semantic integration architectures brings more freedom and promote a weak coupling of the CS, favoring the growing up of the system dependability.

IV. CIGARRA: A NEW ARCHITECTURE

Cigarra is a large-scale project and includes the development of new DCPs, receiving and processing computer systems. The Cigarra Project is a broad concept that can create the new Cemaden observational network management and operation model, based mainly on the management authority decentralization under the DCPs that make up the network. Cigarra uses several modern System Engineering techniques and the Ontology-based Integration represents just one of the techniques presented in the form of fragments in this work. The Fig. 5 represents the high-level implementation of the Ontology-based Integration Architecture.

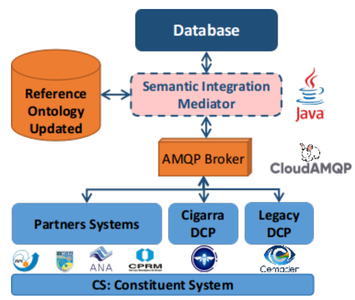

Fig. 5. High-level implementation of the Ontology-based Integration

The first step of the Ontology-based Integration specification process is to obtain the Reference Ontology from the legacy system. This is an important step to identify the integration point and to know what will be the semantic equivalences addressed by the system. The Reference Ontology is represented graphically in Fig. 6.

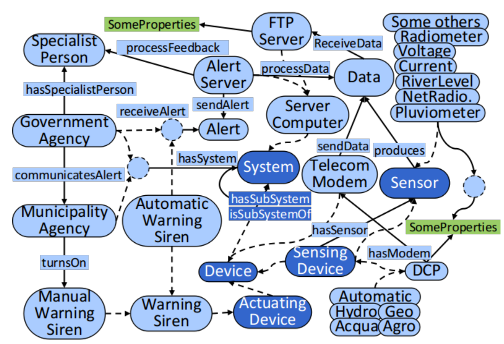

Fig. 6. Cemaden's observational network reference ontology

The complete Reference Ontology file, in Turtle/OWL format, and all documents in this project are available at Github[1].

[1] https://github.com/andreivo/PhD.Ivo/blob/main/ontologies/OR_OBSNetwork.owl

The **Reference Ontology** represents the legacy system and is the main input artifact of the specification process, which, if not available, must be concluded before the changes that support Ontology-based Integration (that generates a new **Reference Ontology Updated**). The complete Reference Ontology Updated is available at Github[2].

Extracted from the Reference Ontology Updated, Fig. 7b illustrates the ontology for the new Data Collection Platforms (DCP) Architecture. The fragment of the ontology shown in Fig. 7b is an evolution of the legacy DCP fragment shown in Fig. 7a.

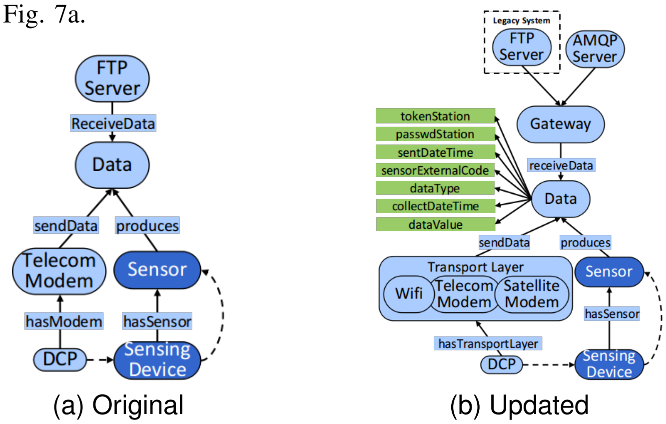

Fig. 7. Generical DCP's Reference Ontology Fragment

As already presented, this project deals with the observational network architecture and the new DCPs generation that will become part of Ontology-based Integration. The Fig. 7 presents an abstract and generical DCP and Fig. 8 shows a Pluviometric DCP, an instance of the generical DCP.

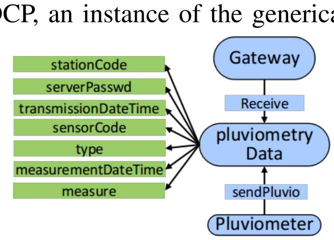

Fig. 8. Pluviometric DCP's Reference Ontology Fragment

In this way, the semantic mapping process of the new Pluviometric DCP is an important artifact to guarantee adherence to the new architecture and will be a fundamental requirement to the Semantic Integration Mediator component. Fig. 9 shows the Pluviometric DCP's semantic mapping, which represents only an architecture and artifacts fragment; in addition to that, the mapping process is interactive and incremental, where the specifications must be refined until the level of detail is sufficient for development.

As shown in Fig. 5, through a Java implementation of the Semantic Integration Mediator, messages are consumed from the CloudAMQP queue, processed, then go through the data semantic validation, and are finally recorded in the database. All code for the Semantic Integration Mediator component is available at GitHub[3].

[2]https://github.com/andreivo/PhD.Ivo/blob/main/ontologies/OR_OBSNetwork_NewArchitecture.owl

[3]https://github.com/andreivo/CicadaProject/tree/main/brokerMediator

Reference Ontology (a) Original	Legacy DCP	Reference Ontology (b) Update
TokenStation	stationCode	tokenStation
PasswdStation	serverPasswd	passwdStation
SentDateTime	transmissionDateTime	sentDateTime
SensorExternalCode	sensorCode	sensorExternalCode
DataType	type	dataType
CollectDateTime	measurementDateTime	collectDateTime
DataValue	measure	dataValue

Fig. 9. Case study: Pluviometric DCP semantic mapping

V. Conclusions

The dynamism of the social vulnerability imposes a great challenge to the mission of Cemaden, making it essential to maintain and expand its observational network to help preserving human lives.

As presented in this work, the current structure of the Cemaden's observational network is based on syntactic rules that generate a strong coupling between its constituent systems, making it difficult for expansion and maintenance.

With a new structure based on the semantic integration of the constituent systems, the observational network architecture becomes more flexible, making the coupling weaker. The weak coupling then enables a safer and failure-free inclusion of new equipment, also allowing the evolution of partner systems with reduced collateral effects on the integration process.

The new architecture proposed in the Cigarra Project demonstrates how the semantic integration process can help mitigate difficulties and increase system dependability, improving the fault tolerance mechanisms which can make the system even more resilient. Finally, the use of ontologies in semantic integration process improves the development of the shared knowledge, which can reduce the cognitive effort to understand the project among the stakeholders.

References

[1] CEMADEN. (2011) Cemaden creation history. [Online]. Available: http://www.cemaden.gov.br/historico-da-criacao-do-cemaden/

[2] L. D. R. LONDE, L. G. MOURA, M. P. COUTINHO, V. MARCHEZINI, and E. SORIANO, "Vulnerability, health and disasters in São Paulo coast (Brazil): challenges for a sustainable development," *Ambiente & Sociedade*, vol. 21, 00 2018.

[3] V. Marchezini, *Dos desastres da natureza à natureza dos desastres*. RiMa Editora, 11 2009, pp. 48–57.

[4] "Decreto nº 7.513 de 1º de julho de 2011: altera o decreto n. 5886 de 6 de setembro de 2006, que aprova a estrutura regional e o quadro demonstrativo dos cargos em comissão e das funções gratificadas do ministério da ciência e tecnologia e dispõe sobre o remanejamento de cargos em comissão." BRASIL. PRESIDÊNCIA DA REPÚBLICA, 2011. [Online]. Available: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/decreto/d7513.htm

[5] CPRM. (2020) Mapa on-line. [Online]. Available: https://geoportal.cprm.gov.br/desastres/

[6] S. Izza, L. Vincent, and P. Burlat, "Ontology-based approach for application integration," in *2005 1st Interoperability of Enterprise Software and Applications*, 02 2005.

[7] R. F. Calhau and R. d. A. Falbo, "An ontology-based approach for semantic integration," in *2010 14th IEEE International Enterprise Distributed Object Computing Conference*, 2010, pp. 111–120.

[8] M. Uschold and M. Grüninger, "Ontologies: Principles, methods and applications," *The Knowledge Engineering Review*, vol. 11, 01 1996.