Improving semantic interoperability of big data for epidemiological surveillance

João L.R. Moreira — Marten van Sinderen — Luís Ferreira Pires — Patricia Dockhorn Costa

University of Twente, 7500 AE Enschede, The Netherlands
{j.luizrebelomoreira, m.vansinderen, l.ferreirapires}@utwente.nl, pdcosta@inf.ufes.br

ABSTRACT: *Future disease outbreaks may spread faster and stronger than recent epidemics (Gates, 2015), such as Zika, Ebola and Influenza. The integration of multiple existing Early Warning Systems (EWS) is a requirement to support disease surveillance in combating infectious disease outbreaks. In this direction, numerous applications have been developed considering big data from diverse sources. However, big data potential can only be exploited if interoperability challenges are addressed. In this paper we discuss a semantic interoperability problem when using data exchanging standards for EWS integration. Particularly, we identify an issue regarding the distinction between the concepts of situation and event in the Emergency Data Exchange Language (EDXL), the OASIS set of standards for disaster management. To cope with interoperability issues we propose an ontology-driven situation-aware approach for the development and integration of EWS. The approach leverages on the OntoEmerge core ontology, in which we incorporate the clear distinction between the concepts of situation and event.*

KEY WORDS: *Disaster risk reduction, disease surveillance, early warning system, semantic interoperability, ontology, situation-awareness, Emergency Data Exchange Language (EDXL)*

1. Introduction

According to the International Strategy for Disaster Reduction (ISDR), a “disaster” is “a situation where serious disruption of the functioning of a community or a society occurs, involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources.” (Othman et al., 2014). Disaster Management (DM), also referred as emergency management or crisis management, addresses an urgent social need. DM is classified as a cyclical process of four phases (Coppola, 2015): (i) mitigation/prevention, for preventing, reducing or eliminating human hazard through risk assessment and resources planning; (ii) preparedness, for prepositioning resources for a disastrous event before it occurs, including long-term mitigation strategies; (iv) response, for responding to a disastrous event through rescue, relief, salvage and immediate damage assessment; and finally, (v) recovery, for returning the affected area and victims' lives back to normality by performing damage assessments, restoration, re-habitation and repair.

Disaster Risk Reduction (DRR) is a new paradigm of DM, which is a systematic approach to analyse and reduce the causal relations of disasters. In epidemic situations, the United Nations Office for Disaster Risk Reduction (UNISDR) is supported by the World Health Organization (WHO) in setting and implementing norms, standards and technical assistance. WHO is the international lead agency for large scale disease surveillance, working together with regional and national government organizations (e.g. CDC and ECDC). Statistics provided by WHO rely on different sources of data: hand-written medical record and ICT-based, call record (tele-health), Electronic Patient Record (EPR) (i.e. Electronic Health Record (EHR)) patient tracking systems and Enterprise Resource Planning (ERP) systems within health facilities. The Sendai framework for DRR[1] describes the requirement of an integrated global warning and response system for epidemics. In particular, it states the need of improvements of the Global Outbreak Alert and Response Network platform[2] through the integration of Early Warning Systems (EWS), which is reinforced by the International Network for Multi-Hazard EWS (IN-MHEWS), which "calls for an integrated and holistic approach to early warnings for multiple hazards and risks tailored to user needs across sectors." (WHO et al., 2015)

Big data is considered in this context because it offers potential to improve efficiency and effectiveness of EWS for disease surveillance. This potential can only be exploited if interoperability challenges can be addressed. In this paper we discuss the semantic interoperability challenges for meaningful integration of big data in EWS for disease surveillance. We propose a framework to tackle these challenges, based on an ontological approach that plays the role of a specification-based situation identification mechanism. It follows the model-driven engineering paradigm, where implementation code, as publisher-subscriber message factories for RESTful services, is generated from model languages. These services are conformant to the EDXL standards from OASIS. This paper is further structured as: Section 2 presents an overview of disease surveillance and EWS, data exchange standards and challenges to achieve semantic interoperability with big data. Section 3 presents the framework architecture to improve semantic interoperability of EWS, the validation plan, the expected benefits and impacts. Finally, Section 4 presents our conclusions and future work.

---

[1] www.unisdr.org/we/coordinate/sendai-framework

[2] www.who.int/ihr/alert_and_response/outbreak-network/en/

2. Background

Traditional surveillance practices rely on individual reporting of disease cases from physicians, veterinarians, infection control practitioners, laboratorians, and medical examiners, being supported by epidemiological and laboratory investigation. However, these are expensive and slow methods, and sometimes not effective for large scale monitoring – when numerous infections need to be analysed with strict time constraints. As technology advances, epidemiological surveillance is leveraged by new sensors, such as human biomonitoring (HBM) and the internet-of-things (IoT). Moreover, using social media and other web-based data sources (e.g. Google news, Wikipedia, blogs) is a promising trend in big data analytics for large scale outbreak monitoring (Amankwah-Amoah, 2015; Bello-Orgaz et al., 2015; Generous et al., 2014; Palen et al., 2010; Sandhu et al., 2016). The combination of existing applications from e-health (e.g. EHR and tele-health) to emerging technologies (e.g. big data, cloud computing, wearable and context-aware computing) is the goal of "disaster e-health", as reported in (Norris et al., 2015).

2.1. *Early Warning Systems*

A disease surveillance system is an EWS that supports the identification of epidemic risks and decision making, combating and minimizing the harms caused by outbreaks. An EWS is a "chain of information communication systems comprising sensor, detection, decision and broker, in the given order, working in conjunction" (Waidyanatha, 2010). The architecture of an EWS is structured as a process of five steps: (i) sensor: perception of elements from the current context; (ii) detection: comprehension of these elements and their relations as a whole; (iii) decision: projection of the wanted future status, i.e. the input for decision selection and decision making; (iv) broker: mediator between the decision and the response components, where knowledge is shared according to message formats; and (v) response: responsible for performing the actions in the field (e.g. evacuation or isolation of a local). Usually the requirements of a response system guide the entire EWS design. Notice that, among the activities of detection, decision and, especially, broker, the exchange of information is fundamental. Therefore, interoperability plays a determinant role in this context.

2.2. *Data exchanging standards*

To address interoperability, international organizations (e.g. HL7 and OASIS) have been producing data exchanging standards. Interoperability can be defined as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged" (IEEE, 1990). A broader definition considers human, social, political, and organizational factors that impact all components of an integrated system. In healthcare ecosystems interoperability could help saving U$ 30 billion per year (GE, 2015). In order to reduce epidemic risks, interoperability must be improved among EWS components, which need to be well integrated for correct and timely response actions. Furthermore, different EWS need to interoperate to achieve a global multi-hazard EWS, allowing the correlation of different types of hazards.

OASIS has an Emergency Management Technical Committee (EM-TC)[3] that maintains a set of specifications, the Emergency Data Exchange Language (EDXL), which enables the "information exchange to advance incident preparedness and response to emergency situations". EDXL is organized in seven categories, each with a specific goal, and together they form an integrated framework for emergency data exchange. The Tracking Emergency Patients (EDXL-TEP) and the Hospital AVailability Exchange (EDXL-HAVE) are crucial for disease surveillance. The Health Level 7 (HL7) set of standards (www.hl7.org) is recognized as the most commonly used hospital messaging standard. HL7 provides a standard messaging and data structure for many healthcare systems. HL7 is intended to be used for internal administration of healthcare facilities, differing from the purpose of EDXL, which focus is "to facilitate emergency information sharing and data exchange across the local, state, tribal, national and non-governmental organizations of different professions that provide emergency response and management services". However, EDXL must be interoperable with HL7 in order to achieve its goals. Collaboration between OASIS and HL7 has resulted in transformation mapping between the standards in order to avoid loss of critical information. This effort supports data exchange between HL7 conformant healthcare applications and EDXL conformant EWS for epidemic situations, when hospitals have to deal with incoming or transferred patients, or even evacuation, with regional coordination and federal support.

Those standardization initiatives are important to achieve common data formats and communication protocols for disaster situations. However, this is not enough to guarantee that the involved actors share messages with same meaning. For example, a patient tracking system may exchange data with an EHR system (in a hospital) following EDXL. The property "vehicle kind" from the "transport type" entity can accept value sets defined by each party. If one generates a message with value "car" meaning an "ambulance" but the other refers to "car" as a "personal car", then they do not share the same semantics. Because of this, the decision making can be affected, leading to erroneous procedures. This is an example of a semantic interoperability problem.

[3] www.oasis-open.org/committees/tc_home.php?wg_abbrev=emergency

2.3. *Semantic interoperability*

At the application level, the interoperability aspects to consider are: (i) coding, i.e. the binary encoding of the messages/streams that carry data (technological interoperability); (ii) formatting, i.e. the packaging of data in a message (syntactic interoperability); (iii) interpretation, i.e. the assignment of meaning to the data (semantic interoperability); and (iv) dialogue, i.e. the process synchronization for the exchange of messages (process interoperability). Coding and formatting are covered by existing standards, but interpretation is only partially covered (Rezaei et al., 2014). Syntactic interoperability refers to data format standards and communication protocols (e.g. EDXL and HL7), enabling systems to communicate to each other. Semantic interoperability refers to the study of meanings, the ability to automatically interpret shared data meaningfully and accurately according to agreed-upon semantics, i.e. a common information exchange reference model. Semantic interoperability focuses in terminology and deals with human interpretation in an unambiguously way, ensuring that the understanding of the information is the same for senders and receivers, ensuring "that the requester and the provider have a common understanding of the "meanings" of the requested services and data" (Heiler, 1995).

A problem found in the foundations of the EDXL standards, which may lead to misunderstandings when exchanging information, is an ambiguous definition of situations, events and incidents, explicit in the glossary of EDXL-TEP and EDXL-SitRep as: "For purposes of this messaging standard, "Situations", "Incidents" and "Events" will be referred to generally as "incidents". (…) An Event is a planned situation (e.g. a parade in Washington DC). "Event" is also used to refer to a situation that has not been formally identified as an incident. (…) EDXL-TEP may refer to any situation whether an incident, event or other occurrence." (OASIS, 2015). The overloaded use of these terms is very common among different sub-domains (e.g. emergency management, local response, hospitals), where an event can refer to a small-scale situation and a situation may refer to a complex event. However, this definition is insufficient to understand and distinguish between situation and event concepts, which are fundamental aspects, and can lead to inconsistent information across systems, different world views of involved parties, and ultimately to erroneous responses.

3. A situation-aware (SA) framework for interoperable EWS

In order to achieve semantic interoperability for meaningful big data integration in EWS for disease surveillance, we propose a framework that supports the development and integration of SA applications. It employs an ontological model-driven engineering approach described in (Moreira et al., 2015). The main idea of this framework is to support the development and modification of applications that participate in the integrated EWS. Consider as an example an EWS to combat Zika virus. A particular application may be responsible for receiving data from HBM sensors attached to patients, providing information of body temperature, heart frequency and Zika symptoms (skin rashes, conjunctivitis, pain, malaise and headache). Another application may be responsible for collecting data from medical records, including whether a patient was bitten by mosquitos, and last lab exams taken. A third application may be a patient tracking system responsible in collecting data from patients being transported in ambulances. A fourth application may be responsible in collecting web-based data and the location of Aedes aegypti infected regions (e.g. HealthMap). A fifth application may be responsible by choosing the appropriate procedures to respond to possible contagions detected. A sixth application may be responsible for sending coordinated messages to regional and federal instances about the recommended procedures. The integration of all these applications must be well-founded to achieve semantic interoperability. Notice that the concept of application here is broad so that an application can even be another integrated EWS, which may have several actors involved.

3.1. *Framework architecture*

The disambiguation method used in this framework is based on ontological analysis, where OntoEmerge (Moreira et al., 2015), the disaster core ontology, serves as a reference model and is well-grounded in the Unified Foundational Ontology (UFO) (Guizzardi et al., 2015). This conceptualization mechanism provides a clear difference between “situation” and “event” concepts, where an “event” is an instance of an “event type”, as a “situation” is instance of a “situation type” (“situoid”). An “event type” can trigger “situation types” and a “situation type” can bring about “event types”. “Event types” have causality and temporal relations among them and can be complex or simple. Furthermore, there is a distinction between an event that triggers the establishment of a situation in the real world and its evidence characterized by data messages. For example, a patient fever (instance of situation type) is triggered by the raise of body temperature up to 37 (a situation creation event) in real world, and this event is evidenced by an event notification, as a data message from a HBM sensor attached to a patient to a server in the hospital, sent through network. Other elements from UFO can improve the semantic expressiveness of situation modelling (e.g. relators and qualities). A software designer can take advantage of these constructs through the combination of OntoUML, the ontological modelling language of UFO for context specification, and the Situation Modelling Language (SML), for the definition of the patterns within a situation to be detected by the EWS. SML plays an important role because it provides specification-based situation identification based on the situation-awareness theory (Wickens, 2008). BPMN can be used for the definition of the process to react to a detected situation of some type. The implementation of the EWS is obtained by an ontological model-driven engineering process, where model transformations are executed from the specification to the realization in some specific technology. Currently, we are working with transformation mappings to a rule-based technology (SCENE, an extension of Drools) for distributed systems (SiNoS) (Costa et al., 2016); and a complex event processing (CEP) solution (an extension of Java ESPER). BPM suites are being analysed for business processes implementation. Among other characteristics, these technologies were chosen because of their potential in high scalable event-data processing.

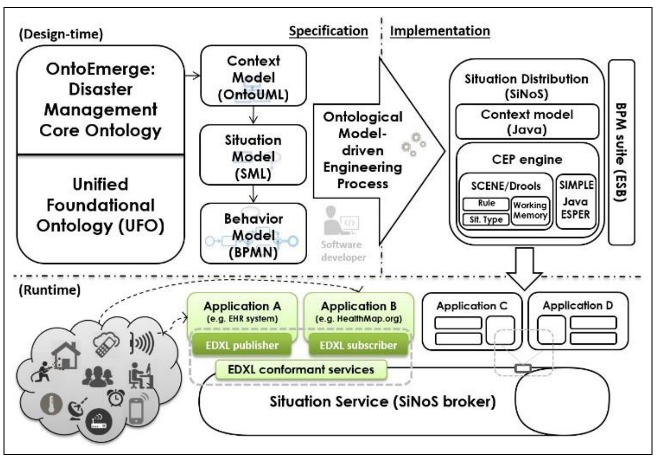

Figure 1. Situation-aware framework for interoperable EWS

At runtime, applications can exchange information as publishers and subscribers of detected situations. The case of two applications implemented with the SCENE platform, they can natively share the situations detected through the SiNoS broker. Applications that do not use this technology can still exchange information by using SOAP and RESTful services exposed through the broker. These services follow the data structures of the EDXL standards, which is possible because of model transformations, from the specification to message formats of the EDXL standards. In particular, regarding the ambiguous definition of situation and event, the mappings distinguish explicitly an event from a situation into the “situation” property of the “PatientType” data structure. This property follows the format of the “SituationType” data structure, which is the “group of elements used to describe the incident associated with the patient". By using this approach it is possible to keep the lightweight property of the EDXL messages while sharpening the semantics of the messages. This is possible because EDXL was designed with a flexible structure that can accommodate different meanings for situation, event and incident. Figure 1 shows the architecture of the framework. On top, the design-time of the application is divided in specification and implementation. On bottom, the runtime of the EWS is illustrated with application A (e.g. an EHR system) as a situation publisher (e.g. provides fever detection service) and B (e.g. HealthMap) as a situation subscriber (e.g. receives fever situation messages), both following the EDXL standards. Applications C and D represent applications implemented with SCENE that are not using EDXL conformant services, but are using other services by sharing their working memories through SiNoS.

3.2. *Validation and intended benefits*

This framework is under development. The mappings for EDXL formats are under construction and it is a research effort that depends on deeper ontological analysis of all EDXL standards. In addition, we expect that this will have an impact on OntoEmerge, which also needs to be aligned to existing healthcare ontologies, such as the BioOntology (www.bioontology.org). Moreover, situation-related concepts in UFO must be harmonized. This will impact the SML metamodel and the editor (to be integrated to a robust modelling tool that enables modelling with OntoUML). The transformation mappings of OntoUML and SML to SCENE and SIMPLE must be improved. To validate the improvement of semantic interoperability with the proposed approach, we plan to use the framework to develop SA applications (the ones described in Section 4.1) as an EWS prototype for Zika and ILI outbreak surveillance. EDXL messages generated from this prototype will be evaluated through interviews with domain experts, e.g. emergency professionals, to check whether knowledge (meaning) sharing could be improved. Situations and related events should be better described, including causality relations between events and between events and situations. As a side effect and a drawback, we expect that the volume of data in EDXL messages will increase because of the additional information (metadata).

4. Conclusions

Developing and integrating applications to create an EWS for global disease surveillance is not a trivial task. Big data needs to be considered to improve the efficiency and effectiveness of EWS, but big data can only be exploited if interoperability challenges are addressed. Syntactic interoperability can be achieved through standardization, such as EDXL from OASIS. However, these standards deal only partially with semantic interoperability issues. Appropriate guidelines are necessary in this domain, especially to bridge the gap between theory and practice. In this paper we propose an approach to address the semantic interoperability issue of EDXL standards regarding the ambiguous definition of situation and event concepts. Therefore, the main contribution of this paper is an ontology-driven framework for meaningful integration of applications within an EWS ecosystem. We argue that semantic interoperability of EWS can be improved with the clear distinction of situation and event concepts reflected into the OntoEmerge ontology. This distinction is represented explicitly in EDXL messages as a result of the model-driven engineering approach, i.e. model transformations.

Current work on this framework includes an ontological analysis of EDXL towards the harmonization of concepts. This effort gives the basis for more accurate model transformations to EDXL. This effort also requires improvement in the OntoEmerge ontology, which needs to be aligned to existing healthcare ontologies. These changes impact the existing modelling tooling support, which must be adapted. Furthermore, we intend to use OntoEmerge as a common understanding mechanism, i.e. a controlled vocabulary (an use-case nomenclature) for the EDXL value sets. We expect to validate our approach by developing new or modifying existing applications that play the role of components of an EWS prototype for global disease surveillance, using data sources related to Zika and ILI cases. Messages generated from situations detected with this EWS will be evaluated through interviews with emergency professionals to measure the improvement of semantic interoperability. The main limitation of our approach is the possible increase of data volume in EDXL messages, which will be assessed to check whether our framework is applicable in real cases.

Acknowledgements

Our thanks to CAPES (PhD scholarship BEX 1046/14-4) and to Patti Iles Aymond, integrant of the OASIS EM-TC and Senior Scientist with IEM (www.iem.com).

References

Amankwah-Amoah, J. (2015). Emerging economies, emerging challenges: Mobilising and capturing value from big data. *Technological Forecasting and Social Change*.

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2015). Social big data: Recent achievements and new challenges. *Information Fusion*.

Coppola, D. P. (2015). Chapter 1 - The Management of Disasters *Introduction to International Disaster Management (Third Edition)*.

Costa, P. D., Almeida, J. P. A., Pereira, I. S. A., Sinderen, M. V., & Ferreira Pires, L. (2016). Rule-Based Support for Situation Management *Fusion Methodologies in Crisis Management*.

Gates, B. (2015). The Next Epidemic — Lessons from Ebola. *The New England Journal of Medicine (NEJM)*.

GE. (2015). GE Healthcare to Connect 500,000 Imaging Machines with Announcement of GE Health Cloud and Apps [Press release]. Retrieved from http://www.genewsroom.com/press-releases/ge-healthcare-connect-500000-imaging-machines-announcement-ge-health-cloud-and-apps

Generous, N., Fairchild, G., Deshpande, A., Del Valle, S. Y., & Priedhorsky, R. (2014). Global Disease Monitoring and Forecasting with Wikipedia. *PLoS Computational Biology*.

Guizzardi, G., Wagner, G., Almeida, J. P. A., & Guizzardi, R. S. S. (2015). Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story. *Journal of Applied Ontology*.

Heiler, S. (1995). Semantic interoperability. *ACM Comput. Surv., 27*(2), 271-273. doi:10.1145/210376.210392

IEEE. (1990). IEEE Standard Glossary of Software Engineering Terminology. New York, NY: IEEE Std 610.12-1990.

Moreira, J. L. R., Ferreira Pires, L., Van Sinderen, M., & Costa, P. D. (2015). Towards ontology-driven situation-aware disaster management. *Journal of Applied Ontology*.

Norris, A. C., Martinez, S., Labaka, L., Madanian, S., Gonzalez, J. J., & Parry, D. (2015). Disaster E-health: A new paradigm for collaborative healthcare in disasters. *ISCRAM 2015 Conference Proceedings - 12th International Conference on Information Systems for Crisis Response and Management*.

OASIS. (2015). Emergency Data Exchange Language (EDXL) Tracking of Emergency Patients (TEP) Version 1.1 - EDXL-TEP *Emergency Management TC*.

Othman, S. H., Beydoun, G., & Sugumaran, V. (2014). Development and validation of a Disaster Management Metamodel (DMM). *Information Processing and Management*.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010). A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. *Proceedings of the 2010 ACMBCS Visions of Computer Science Conference*.

Rezaei, R., Chiew, T. K., & Lee, S. P. (2014). An interoperability model for ultra large scale systems. *Advances in Engineering Software*.

Sandhu, R., Gill, H. K., & Sood, S. K. (2016). Smart monitoring and controlling of Pandemic Influenza A (H1N1) using Social Network Analysis and cloud computing. *Journal of Computational Science*.

Waidyanatha, N. (2010). Towards a typology of integrated functional early warning systems. *International Journal of Critical Infrastructures*.

WHO, WMO, UNDP, UNESCO-IOC, UNESCAP, UNISDR, . . . GIZ. (2015). *International Network for Multi-Hazard Early Warning Systems (IN-MHEWS)*.

Wickens, C. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *The Journal of the Human Factors*.