NWPOntology: A well-founded Ontology for the Numerical Weather Prediction Data Domain

Tiago Duarte Neves[1], Ivanildo Babosa[1], Kelli de Faria Cordeiro[1,2]

[1] Military Institute of Engineering (IME), Rio de Janeiro – RJ 22290-270, Brazil

[2] Center of Naval System Analysis of Brazilian Navy, Rio de Janeiro – RJ 20091-000, Brazil

tiago.duarte@ime.eb.br, ivanildo@ime.eb.br, kelli@marinha.mil.br

Abstract. *Weather forecasting is an important application of Science to plan daily activities, being used in agriculture, air traffic, sea, forestry, severe weather alerts, military operations and enterprise systems. The weather forecast information is provided by several systems with different formats, parameters and with a large number of heterogeneous concepts used in different organizations. To minimize this problem, the number of these heterogeneous concepts must be reduced in order to provide uniform access for better decision making. This article describes the construction of a well-founded ontology for the Numerical Weather Prediction data domain (NWPOntology), developed from philosophically based foundations of the Unified Foundational Ontology as a solution to promote the semantic enrichment of data from meteorological observations, aiming to reduce conceptual ambiguities.*

1. Introduction

Existing weather forecast centers focus on presenting data to the forecaster supporting some important decisions. Most of the data analysis and assembling key decisions in a forecasting policy takes place in the minds of analysts. This is usually transformed into forecast products finished in a simple text editor with large amounts of complex, semi-structured and multivariate data, having to interpret them in a short time. Therefore, one of the biggest challenges in geospatial research is to generate accurate data.

In Meteorology, it is necessary ensure data consistency in the most efficient way. Inconsistencies and variations in the data forecasting series may occur due to several reasons, such as, human error in data acquisition, inadequate maintenance and calibration of the measuring instrument, erroneous data processing, random variations expected for any weather phenomenon and changes in the surrounding environment. In this context, the weather forecasting centers may use different syntaxes and terminologies for describing information as well as the same term may have different meanings and interpretations in different organizations.

For example, in the sentence "Maps of daily temperature and precipitation are produced," an expert would recognize that the observation is "temperature" but could not determine the details related to the temperature concept (atmospheric temperature, sea surface temperature, etc.). He needs to ask the data provider to get more details [Masmoudi et al. 2020]. Additionally, in different disciplines, the same term may correspond to various meanings, in one hand. For example, the term "environment" is defined as "the biological and abiotic elements surrounding an individual organism," in the biological domain. However, it refers to "all the natural components of the Earth (air, water, oils, etc.)" [Sauvé et al. 2016].

On the other hand, various terms may correspond to the same meaning. For instance, the Observatory of Sahara and Sahel (OSS) may use the word "rainfall" for the same real-world feature that usually refers to "precipitation" in other sources. The variety of terms complicates the work of emergency responders who should be familiar with the terms used in each discipline [Masmoudi et al. 2020]. Another situation that occurs in an organization might use the term "river", while another might use the term "lake" to describe the concept "water area" [Khantong and Ahmad 2019]. Consequently, although the meanings of the concepts for describing information from these organizations are similar, stakeholders and computational systems in different organizations cannot understand the meanings.

Semantic Sensor Web (SSW), SWEET and Next Generation Network Enabled Weather (NNEW) are ontologies designed to represent the domain of weather data (to be detailed in the Section2), but they did not emphasize numerical weather prediction, which is essential for an area that is constantly evolving. To provide conceptual methodological support for the exchange of knowledge on Numerical Weather Prediction Data, NWPOntology is being developed, a domain ontology for numerical weather prediction data based on Ontology of Unified Foundational Ontology (UFO) [Guizzardi 2005] to support a future decision-making system. The adoption of NWPOntology could improve the safety of marine navigation by users as well as the semantic accuracy of that data by Command and Control applications, e.g., Search and Rescue operations.

This document introduces NWPOntology as follows: Section 2 reviews some related work from ontology in weather data domain; Section 3 details NWPOntology; and finally, Section 4 presents final observations and topics for further investigation.

2. Related Work

There exists a huge number of sensor networks around the globe designed to monitor a large set of different phenomena, creating a vast amount of data. Making them available demands to structure such data as well as to allow the interoperability between different sensor networks. Semantic Sensor Web (SSW) is an approach that builds upon Sensor Web Enablement (SWE) [Reed et al. 2007] and Semantic Web activities by the W3C [W3C 2020] which aims at annotating sensor data with semantic metadata to increase interoperability and to provide contextual information essential for situational knowledge [Sheth et al. 2008]. Semantic metadata includes spatial, temporal, and thematic data.

The SSW ontology does not support forecast values. However, it would be possible to extend the SSW ontology to implement them. In the context of NWPOntology, the concept *Observations*, as well as its subconcepts, could be reused; however, some additional subconcepts should to be added. The Basic (WGS84 lat/long) for location [Brickley 2004] and OWLTime [Hobbs and Pan 2006] can be used by the NWPOntology without the use of the SSW ontology.

Another approach towards a semantically-enriched sensor network, based on SWE, is an OWL 2 ontology created by the W3C Semantic Sensor Network Incubator group (SSN-XG) which is referred to as the SSN Ontology [Compton et al. 2012]. The goal of this ontology is to simplify managing, querying, and combining sensors and observation data from different sensor networks. Different perspectives can be used to view the knowledge base, including: the sensor perspective (which sensors are available; what and how do they sense), the observation perspective (focusing on observations and related metadata), the system perspective (systems of sensors), and the property perspective (properties of physical phenomena and how they are sensed).

SWEET is a set of more than 200 ontologies comprising about 6000 concepts [Savić et al. 2019]. The initial version dates back to 2003 and based on DAML+OIL [Connolly et al. 2001, Raskin and Pan 2003]. Works regarding environmental applications include integrating volcanic and atmospheric data in the context of volcanic eruptions [Fox et al. 2007], an ontology of earth's crust fractures [Zhong et al. 2009] and an extension of SWEET by climate and forecast terms [Ramachandran and Raskin 2006].

In the context of weather forecasting, SWEET ontologies are a well-qualified approach for reusing them in a climate ontology. Time definitions, and geographical position specifications are supported. However, SWEET comes with some disadvantages, which ultimately led to the decision not to use them in the implementation of NWPOntology to support a future decision-making system: it does not currently cover future events, as well as does not cover all weather elements relevant to numerical weather prediction data, e.g., numerical weather prediction models.

The Next Generation Network Enabled Weather (NNEW) ontology is designed to provide a comprehensive view on the weather across the country, built from thousands of single weather observations. One of the tasks foreseen for the system was the development of ontologies for the meteorological domain, mainly for the reduction of delays related to the climate in the airspace of the USA to half of its current magnitude. Recently, approximately 70 percent of all air traffic delays are attributable to weather [Blasch et al. 2019].

This ontology was built with respect to SWEET ontologies to map weather phenomena. Extensions to SWEET include additional weather phenomena and concepts and relations that lead to the 4-D Wx Data Cube (Four Dimensional Weather Data Cube) which uses time as the fourth dimension for the "location" of weather observations. Just like SWEET, NNEW ontology does not address the context of weather predictions. This was the reason for not using this ontology.

The Section 2 discusses no existing foundational ontology covers the domain of numerical weather prediction data, in a way suitable for using it as a starting point for NWPOntology. Thus, the creation of NWPOntology was based on UFO conceptual requirements, reusing of existing meteorological ontology concepts (see Section3). The ability to clearly express concepts of the real-world, reducing conceptual ambiguities, leads the scientific community, as well as conceptual modeling professionals, to consider UFO as an important resource to model the domain ontologies.

3. NWPOntology: A well-founded Ontology for the Numerical Weather Prediction Data Domain

Detailed surveys on ontology development methodologies can be found in the works of [Corcho et al. 2003, Iqbal et al. 2013]. To construct NWPOntology the methodology used was the Systematic Approach to Build Ontologies (SABiO) [Falbo 2014]. SABiO methodology having been created to support the development of domain reference ontologies and assumes that domain ontologies must be developed based on foundational ontology. The ontology development process comprises five main phases: Purpose Identification and Requirements Specification; Capture and Formalization of Ontology; Project and Test.

In the Purpose Identification and Requirements Specification phase, it was defined that NWPOntology aims to serve as a shared conceptualization to resolve the semantic heterogeneity caused by divergent interpretations of data according to the different contexts in which they are used, and the green part of the model will support an analytical environment of historical events. Its main objective is the modeling and implementation of a domain ontology for numerical weather prediction data (NWPOntology), as a solution to reduce the conceptual ambiguities of the weather forecast data existing in the various forecast centers, promoting semantic enrichment meteorological observation data. In this way, the information can be observed, analyzed and disseminated by weather reports, assisting in the decision making process.

The focus of the Capture and Formalization phase is to capture the conceptualization of the discourse universe. Relevant concepts and relationships must be identified and organized. This phase begins with conceptual modeling using highly expressive languages to create strongly axiomatized ontologies that are as close as possible to the domain's ideal ontology. The focus of these languages should be on the adequacy of the representation of the main concepts and relationships in the domain using taxonomies.

The NWPOntology fragment presented in this article is based mainly on the taxonomy of meteorological data terms proposed in [Da Cruz et al. 2015]. Other concepts found in NWPOntology were obtained by reusing the concepts of previous knowledge resources, such as GeoSPARQL [OGC 2020], INSPIRE [INSPIRE 2019], ISO 19156:2011 [ISO 2011], GEneral Multilingual Environmental Thesaurus [GEMET 2020] and World Meteorological Organization Glossary [WMO 2020]. In formalization, what is intended is to explicitly represent the conceptualization captured in the previous stage with the use of informal axioms written in a formal language, with OntoUML being chosen to play this role.

NWPOntology model illustrated in Figure 1, making the distinction between different types of entities explicit in the ontological sense. The UFO-B stereotypes present in our modeling (the green part) were adapted from [Botti Benevides et al. 2019].

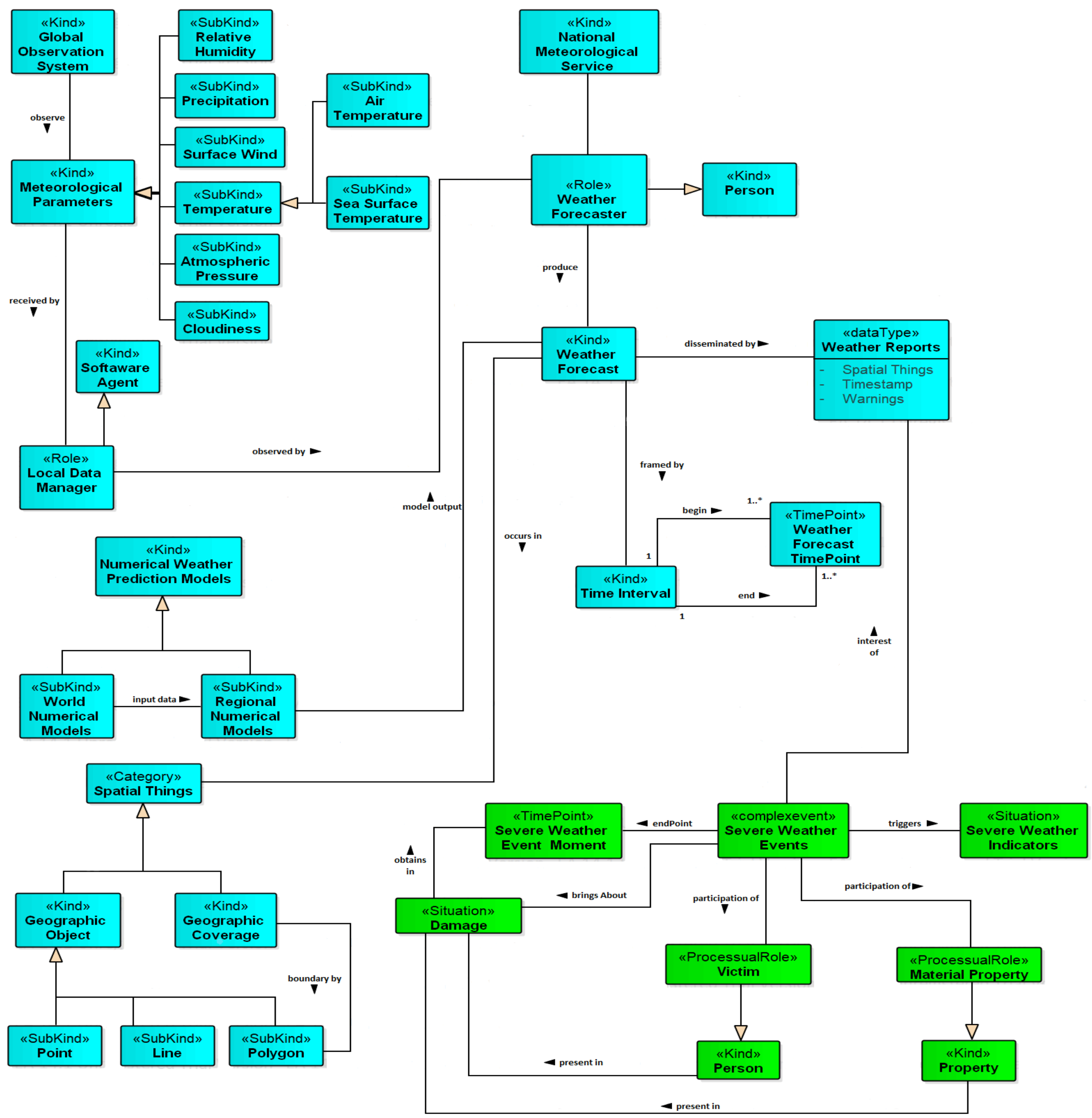

Figure 1. NWPOntology model based on UFO-A and UFO-B

The meteorological parameters, modeled as the <kind> stereotype, is used to represent rigid concepts, that provide an identity principle for their instances and do not require a relational dependency. They are basically divided into variables, e.g., values of relative humidity, precipitation, surface wind, air temperature and sea surface temperature, atmospheric pressure, and cloudiness. Such concepts are modeled by the <subkind> stereotype, so they have their own identities and inherit the properties of the meteorological parameters class. In this case, the data subtypes were mapped as a <subkind> sortal, precisely because they are unique, but at the same time inherit characteristics in common.

Furthermore, according to our model, they are observed by the Global Observation Systems sensors, that was modeled by the <kind> stereotype, which is mapped by the formal relationship created between the Global Observation System and the observed meteorological parameters. These data are received at the National Meteorological Services (modeled by the <kind> stereotype) and the data is transformed into a graphical visualization, through the Local Data Managers (LDM). The LDM was modeled by the <role> stereotype, because is a construct used to represent anti-rigid specializations of identity providers, in this case the Software Agent, that was modeled by the <kind> stereotype.

Meteorological data transformed into a graphical visualization is observed by the Weather Forecasters and compared with the data generated by the outputs of the Regional Numerical Models. Weather forecasters, modeled by the <role> stereotype, which has its identity principle inherit by the National Meteorological Services class. Regional Numerical Models, modeled by the <subkind> stereotype, for they have their own identities and inherit the properties of the Numerical Weather Prediction Models class, modeled by the <kind> stereotype.

Global Numerical Models, modeled by the <subkind> stereotype, for they have their own identities and inherit the properties of the Numerical Weather Prediction models class, input the meteorological parameters data for the Regional Numerical models. If there are no discrepancies between the outputs of the numerical models and the meteorological data of the LDM, the Weather Forecaster produces the Weather Forecast, modeled by the <kind> stereotype to be disseminated as the Weather Reports, modeled by the <datatype> stereotype. This stereotype represents data whose members are numerical values belonging to a given set of possible values. A concept represented as a <datatype> is a theoretical representation of a conceptual space and the restrictions imposed by its respective geometric structures [Guizzardi and Wagner et al. 2004]. A value does not have an identity, so two occurrences of the same value cannot be differentiated [Guizzardi 2005], that occurred in a certain location and during a certain time.

Severe Weather Events refers to any dangerous meteorological phenomena with the potential to cause damage, serious social disruption, or loss of human life. Types of Severe Weather Events vary depending on the latitude, altitude, topography and atmospheric conditions. High winds, hail, excessive precipitation and wildfires are forms and effects of severe weather, as are thunderstorms, downbursts, tornadoes, waterspouts, tropical cyclones, and extratropical cyclones, and should be released through weather reports as soon as possible to users [WMO 2020].

Severe Weather Events consist of a complex event that brings about the victim or property to a damage situation, represented at the time (Severe Weather Event Moment), when the victim or property suffers that damage. A Severe Weather Events only occurs if it is triggered by a situation of significant change in atmospheric parameters (Severe Weather Indicators). An example of a Severe Weather Event would be a Tropical Cyclone occurring in a given location, causing material damage and death to people.

4. Final considerations

The need to reduce the conceptual ambiguities of meteorological data, which make up the Global Observation System, motivated the construction of NWPOntology, aiming at an open and generic approach to interoperability between the different systems for the correct decision making. To complete the development of the ontology it is still necessary to have a dynamic verification and validation of the behavior of the operational ontology, in a finite set of test cases against the expected behavior in relation to Competence Questions (CQs). Investigations and experiments in

NWPOntology are still needed, mainly with regard to studies on UFO-B event stereotypes, as the incorporation of new entities is necessary, such as material damage caused by serious climatic events in regions of the world potentially affected. Therefore, future research can benefit from the ontological concepts identified in this research.

References

Blasch, E. P., Insaurralde, C. C., Costa, P., De Waal, A. and De Villiers, J. P. (2019). Uncertainty Ontology for Veracity and Relevance. In *2019 22th International Conference on Information Fusion (FUSION)* (pp. 1-8). IEEE.

Botti Benevides, A., Bourguet, J. R., Guizzardi, G., Peñaloza, R. and Almeida, J. P. A. (2019). Representing a reference foundational ontology of events in SROIQ. *Applied Ontology*, *14*(3), 293-334.

Brickley D., editor. (2004). Basic Geo (WGS84 lat/long) Vocabulary. http://www.w3.org/2003/01/geo/. Accessed: 2020-03-08.

Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S. and Huang, V. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, *17*, 25-32.

Connolly, D., van Harmelen, F., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F. and Stein, L. A. (2001). DAML+ OIL (2001) Reference Description. W3C Note, 18 December (2001). Accessed: 2020-03-05. Latest version available at http://www.w3.org/TR/daml+oil-reference.

Corcho, O., Fernández-López, M., Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point? Data Knowl Eng 46(1):41–64.

Da Cruz, S. et al. (2015). Gestão Semântica de Dados Meteorológicos Apoiados Por Ontologias de Proveniência. In: Anais do IX Brazilian e-Science Workshop. SBC, p. 61-70

Falbo, R.D.A. (2014). SABiO: Systematic Approach for Building Ontologies. In *ONTO. COM/ODISE@ FOIS*.

Fox, P., McGuinness, D., Raskin, R. and Sinha, K. (2007). A volcano erupts: semantically mediated integration of heterogeneous volcanic and atmospheric data. In Proceedings of the ACM first workshop on CyberInfrastructure: information management in eScience (pp. 1-6).

GEMET-Concepts (2020). General European Multilingual Environmental Thesaurus, Version. 4.1.4 - Rapport under Contract of the European Environment Agency.

Guizzardi, G., Wagner, G., Sinderen, *et al.* (2004). "An Ontologically Well-Founded Profile for UML Conceptual Models", In: **Proceedings of the 16th International. Conference on Advances in Information Systems Engineering**, pp. 112-126.

Guizzardi, G. (2005). Ontological Foundations for Structural Conceptual Models. PhD Thesis, University of Twente, Netherlands.

Hobbs, J. R. and Pan, F. (2006). Time ontology in OWL. W3C working draft, 27 September 2006. *World Wide Web Consortium*.

INSPIRE (2019). Infrastructure for Spatial Information in Europe. Available: https://inspire.ec.europa.eu/. Accessed: 2020-03-12.

Iqbal, R., Murad, M.A.A., Mustapha, A., Sharef, N.M. (2013). An analysis of ontology engineering methodologies: a literature review. Res J Appl Sci Eng Technol 6(16):2993–3000.

ISO (2011). ISO 19156:2011 - Geographic information - Observations and measurements.

Khantong, S. and Ahmad, M. N. (2019). An Ontology for Sharing and Managing Information in Disaster Response: In Flood Response Usage Scenarios. *Journal on Data Semantics*, 1-14.

Masmoudi, M., Karray, M. H., Lamine, S. B. A. B., Zghal, H. B. and Archimede, B. (2020). MEMOn: Modular Environmental Monitoring Ontology to link heterogeneous Earth observed data. Environmental Modelling & Software, 124, 104581.

OGC (2020). Open Geospatial Consortium | OGC®. http://www.opengeospatial.org/. Accessed: 2020-03-12.

Ramachandran, R., Graves, S. and Raskin, R. (2006). Ontology re-engineering use case: Extending SWEET to map climate and forecasting vocabulary terms. In AGU Spring Meeting Abstracts.

Raskin, R. and Pan, M. (2003). Semantic web for earth and environmental terminology (sweet). In Proc. of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data (Vol. 25).

Reed, C., Botts, M., Davidson, J. and Percivall, G. (2007). OGC® Sensor Web Enablement: Overview and High Level Achhitecture. In *2007 IEEE Autotestcon* (pp. 372-380). IEEE.

Sauvé, S., Bernard, S. and Sloan, P. (2016). Environmental sciences, sustainable development and circular economy: Alternative concepts for trans-disciplinary research, In Environmental Development, Volume 17, 48-56, ISSN 2211-4645.

Savić, M., Ivanović, M. and Jain, L. C. (2019). Analysis of ontology networks. In *Complex Networks in Software, Knowledge, and Social Systems* (pp. 143-175). Springer, Cham.

Sheth, A., Henson, C., and Sahoo, S. S. (2008). Semantic Sensor Web. *IEEE Internet computing*, *12*(4), 78-83.

WMO. (2020). *WMO Sea-ice Nomenclature: Terminology, Codes and Illustrated Glossary*. Secretariat of the World Meteorological Organization.

W3C (2020). Semantic Sensor Network Incubator Group. http://www.w3.org/2005/Incubator/ssn/. Accessed: 2020-03-12.

Zhong, J., Aydina, A. and McGuinness, D. L. (2009). Ontology of fractures. Journal of Structural Geology, 31(3), 251-259.