

# Using Well-Founded Provenance Ontologies to Query Meteorological Data

Thiago Silva Barbosa<sup>1</sup>, Ednaldo O. Santos<sup>1</sup>, Gustavo B. Lyra<sup>1</sup>,  
and Sérgio Manuel Serra da Cruz<sup>1,2,3(✉)</sup>

<sup>1</sup> UFRRJ – Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, Brazil  
{thiago, ednaldo, gblyra, serra}@ufrrj.br

<sup>2</sup> PPGMMC/UFRRJ – Programa de Pós Graduação Modelagem  
Matemática e Computacional, Seropédica, RJ, Brazil

<sup>3</sup> PET-SI/UFRRJ – Programa de Educação Tutorial - Sistemas de Informação,  
Seropédica, RJ, Brazil

**Abstract.** The analysis of increasing flow of data about Tropical rainfall is a big challenge faced by meteorologists. This work presents an approach to pre-process, organize and query high quality meteorological data. Thus, we present a semantic approach that uses well-founded ontologies that help meteorologists to develop SPARQL queries that navigate over high quality data and provenance metadata collected during the execution meteorological in silico experiments.

**Keywords:** Provenance · Foundational ontology · Meteorology

## 1 Introduction

There is great interest in determining the periods and the probability of occurrence of extreme hydrometeorological events so as to mitigate possible associated risks to citizens and agribusiness. Briefly, meteorological data flows from lots of sensors through heterogeneous apparatus to scientists' databases where they perform statistics, analytics to tune mathematical models to study the occurrences extreme events. Therefore, in this work, we present an approach that uses well-founded ontologies [1, 3, 8] and provenance management techniques to aid researchers to investigate the cause of erroneous values detected at any point of the pre-processing chain and to query high quality meteorological.

## 2 Materials and Methods

**Meteorological Data and Pre-processors** - Daily raw rainfall data were obtained from 75 weather stations geographically scattered in the southeast region of Rio de Janeiro State, Brazil, one of the regions subject to the occurrence of extreme rainfall events. The datasets are part of long meteorological series (bigger than 20 years since 1960). The series were extracted over the Web from FAO and HidroWeb systems [4] by a Web framework named “Metoro”, previously developed by our research group [2], which uses several Vistrails workflows as chains of pre-processors to generate higher

quality curated meteorological data. The pre-processors checks: high–low extreme daily values, internal consistency, temporal and spatial outliers, missing and erroneous data. The framework allows meteorologists to rectify gap data and annotate datasets with provenance to reduce error propagation on long-term meteorological investigations. Besides, the framework also generates a structured relational repository of high quality meteorological data. The quality of data in the repository generated was evaluated by Precinoto et al. (2013) [5]. However, despite the computations, data are still faulty and presents some semantic inconsistencies. Thus, in order to reduce the semantic gaps, we developed well founded provenance ontology to annotate meteorological data of the repository.

**Well-Founded Ontologies** - In this work, we have used the ontologically well-founded UML modeling profile named OntoUML presented by Guizzardi and Halpin [3] to develop well-founded ontologies. This profile comprises a number of stereotyped classes and relations implanting a metamodel that reflect the structure and axiomatization of a foundational and domain independent ontology named Unified Foundation Ontology (UFO). We also used the Open proVenance Ontology (OvO) [1] which is based in three other theories: the lifecycle of scientific experiments, presented by Mattoso et al. [6], PROV-O and PROV-DM specifications and UFO itself. OvO's concepts are modeled as UML profile because of the widespread understanding of classes and relations and their suitability. OvO was developed as a set of three sub-ontologies: (i) *in silico* scientific experiment sub-ontology, (ii) experiment composition sub-ontology, (iii) experiment execution sub-ontology. The sub-ontologies complement each other; they are connected by relations between their concepts as well as by formal axioms.

### 3 Meteorology Ontology and WebOntology Query Tool

*Meteoro* is an application ontology that maps the concepts of (i) the pre-processing steps of raw meteorological data into curated data; (ii) provenance metadata about data transformations executed by the pre-processors and; (iii) the characteristics about the *in silico* experiments performed by the meteorologists. It makes these concepts explicit, extends the OvO to that domain, besides reuses the concepts of provenance in large scale scientific experiments described by Cruz et al. [1, 8]. *Meteoro*, like OVO, was designed using OLED (OntoUML Lightweight Editor) [3], it is an editor for OntoUML, aimed to provide a simple, lightweight and integrated set of features such as model editing, syntax verification, instances simulation via Alloy, anti-pattern management and transformations to OWL. In other words, *Meteoro* is first modeled in an ontologically well-founded language that explicitly commits to fundamental ontological distinctions in their metamodels comprising type such as: Rigid (*Kinds* and *subKinds*), Anti-Rigid (*Phases* and *Roles*) and Semi-Rigid (*Mixins*). After that, it can be converted to another language that supports inferences and reasonings.

***Meteoro* Ontology** - To be computed, the ontology has to be codified into another language that supports automated inferences. Besides, it must consider legacy applications and other relevant requirements such as reasonable computational efficiency

and compatibility with Semantic Web standards. Thus, we transformed *Meteoro* from OLEd to OWL taking advantage of the Protégé editor. The codification of well-founded ontologies to OWL is complex. The mappings between two radically different languages need for customizations to represent each domain element. During the execution of this work OLEd was still under development; thus we used two rounds of mapping. As the first round, we used the mapping rules defined by Zamborlini et al. [7]. As the second round of mapping, we used rules to match the concepts of the ontology to the relations of the meteorological repository. This approach allows relational databases to offer their contents as virtual RDF graphs without the replication of the RDB in RDF triples. Besides, it permits meteorologists to develop SPARQL queries and navigate over meteorological data and provenance metadata through the concepts of the ontology.

**WebOntology Tool** - We have noticed that it was not trivial for meteorologists to create SPARQL queries that involve meteorological data, provenance metadata and also ontology classes. Thus, we developed a simple web-based graphical query tool named WebOntology that uses the *Meteoro* ontology to assist meteorologists with respect to the process of query formulation over the meteorological repositories. There are two main functionalities that we considered important to be mentioned: (i) *Manage Queries*: It aims to reduce the researcher's (re)work. It allows them to create, execute, delete and update SPARQL queries over the data repository; (ii) *SPARQL EasyBuilder*: It lets meteorologists create simple queries even without knowing the syntax of the language. Therefore, it allows users to navigate through the concepts and properties and graphically develop simple queries by selecting features like ontology class, object, properties and values to be searched.

## 4 Conclusion

This work presented an approach to help meteorologist to manage curated data about Tropical rainfall. Our proposal incorporates well-founded ontologies, provenance and Semantic Web standards to recover high quality meteorological data annotated with provenance metadata generated during early stages of data transformation.

**Acknowledgements.** We are grateful by the financial support provided by FAPERJ (E-26/112.588/2012 and E-26/110.928/2013) and FNDE-MEC-SeSU.

## References

1. Cruz, S.M.S, Campos, M.L.M., Mattoso, M.: A foundational ontology to support scientific experiments (2012). [ceur-ws.org/Vol-728/paper6.pdf](http://ceur-ws.org/Vol-728/paper6.pdf)
2. Lemos Filho, G.R., et al.: Assimilação, Controle de Qualidade e Análise de Dados de Meteorológicos Apoiados por Proveniência. In: VII Brazilian E-science Workshop (2013)
3. Guizzardi, G., Halpin, T.: Ontological foundations for conceptual modeling. *Appl. Ontol.* **3**, 91–110 (2008)

4. HidroWeb: Sistemas de Informação Hidrológicas (2014). <http://hidroweb.ana.gov.br/>
5. Precinoto, R.S., et al.: Aplicação de Regressão Linear Múltipla para Preenchimento de Falhas de Dados Pluviométricos no Estado do Rio de Janeiro. In: Anais XVII SBMET (2012)
6. Mattoso, M., et al.: Towards supporting the life cycle of large scale scientific experiment. Int. J. Bus. Process Integr. Manage. **5**(1), 79–92 (2010)
7. Zamborlini, V., Gonçalves, B., Guizzardi, G.: Codification and application of a well-founded heart-ECG ontology (2011). [http://www.inf.ufes.br/~gguizzardi/camera-ready\\_paper48363.pdf](http://www.inf.ufes.br/~gguizzardi/camera-ready_paper48363.pdf)
8. Cruz, S.M.S.: Uma Estratégia de Apoio à Gerência de Dados De Proveniência em Experimentos Científicos. Ph.D. Thesis, Federal University of Rio de Janeiro - COPPE, Brazil (2011)