# Development of an Ontology-Based Approach to Spatio-Temporal Data Analysis for Forest-Environment Interactions: Extended Abstract*

Kingsley Wiafe-Kwakye*,†, Torsten Hahmann† and Kate Beard†

*School of Computing and Information Science, University of Maine*

Abstract

In response to rapidly changing climate, management of future forests requires better analytical tools to understand the complex interactions within forest ecosystems. While increasingly fine-grained data about forests, individual trees, and closely connected biotic and abiotic factors are collected, their successful spatial, temporal and semantic integration are crucial for holistic forest analytics.

We address this integration challenge by developing a novel ontology-based GIS—the Digital Forest—for supporting forest resource management. Two core advances that enable the Digital Forest are presented here: (1) critical and novel ontologies that connect the disparate forest knowledge into a large semantically-explicit geospatial knowledge graph, and (2) a novel approach for using machine learning techniques in conjunction with the ontologies to spot and summarize hidden associations. The tool's capabilities for semantically querying, exploring, visualizing, and analyzing geospatial forest knowledge are demonstrated using data from Northern New England.

**Keywords**

Ontology, forestry, data integration, environment

## 1. Motivation

Nearly every environment on Earth, ranging from freezing Arctic environments to boiling vents deep beneath the ocean's surface, support life. Each ecosystem provides resources and limitations that dictate the kind and distribution of the species that live there [1]. In addition to being shaped by their environment, organisms also alter their environment at varying scales as a result of their ongoing interactions with their environment and with one another. Forests are a major component of the Earth's ecosystems that occupy an estimated area of 4.06 billion hectares which is approximately 31% of the total land area [2]. Forests, being a significant component of the Earth's ecosystems, provide essential ecosystem services, including regulation

*FOIS 2023 Early Career Symposium (ECS), held at FOIS 2023, co-located with 9th Joint Ontology Workshops (JOWO 2023), 19-20 July, 2023, Sherbrooke, Québec, Canada*

*Corresponding author.

†These authors contributed equally.

kingsley.wiafekwakye@maine.edu (K. Wiafe-Kwakye); torsten.hahmann@maine.edu (T. Hahmann); kate.beard@maine.edu (K. Beard)

https://umaine.edu/scis/people/torsten-hahmann/ (T. Hahmann); https://umaine.edu/scis/people/kate-beard-tisdale/ (K. Beard)

0009-0007-3741-3721 (K. Wiafe-Kwakye); 0000-0002-5331-5052 (T. Hahmann); 0000-0002-7703-0270 (K. Beard)

of water quality, flood mitigation, erosion control, and storing carbon ([2], [3], [4], [5]).

Forests are complex environmental systems as a result of the interactions between several biotic and abiotic factors that contribute to forest ecosystems, such as interactions among trees and interactions between trees and rapidly changing climatic conditions. These interactions impact the ecology of forest systems thereby greatly influencing the provision of ecosystem services now and in the future [6]. Better understanding of the forest ecosystem will help make more informed decisions on forests management. This deeper understanding however, requires both community-based and large-scale research in forested ecosystems over large geographical and temporal extents which requires the synthesis of a large amount of data.

Modern forestry surveys are producing a large amount of data due to advanced data collection and processing technologies, as opposed to traditional methods that rely on sampling and visual techniques. While this creates an opportunity for large-scale forest research, there is a challenge of integrating these large and diverse datasets into a single system due to the fact that ecological data covers a broad spectrum of disciplines, data formats, structures, and semantic concepts. There is therefore a concrete need to overcome the integration challenges to allow for the synthesis of different kinds of environmental and forest inventory data into an integrated knowledge base.

## 2. Research Questions

1. What semantics are needed to describe and reason about nuances of spatio-temporal aggregation, common forest domain concepts, and environmental preferences of species to help integrate, interpret and analyze forestry data?
2. How can ontologies be used in conjunction with machine learning techniques to facilitate interpretable exploration and analysis of forestry data?

## 3. Objective(s)

The overarching objective is to identify and design ontologies for formal representation, interpretation and integration of forest data and to devise novel approaches that leverage synergies between these ontologies and machine learning approaches to improve integrated forestry data analysis.

### 3.1. Specific Objectives

1. To develop an ontology pattern for resolving semantic ambiguity regarding spatial and/or temporally aggregated quantities
2. To formalize the semantics of forest-related concepts as an OWL ontology to provide a common framework for representing forest-related concepts
3. To develop a formal semantic representation that connects tree species to their preferred environmental conditions (preference habitat)
4. To deploy the ontologies together with the data and execute test queries to evaluate their effectiveness for expressing and answering competency questions from the forestry domain.

5. To demonstrate how ontology embeddings and similarity measures could be used in machine learning to allow for hybrid logical-statistical queries and analysis to improve our understanding of forest ecosystem interactions.

# 4. Research Methodology

The overall approach used is to identify and, if necessary amend or design new, ontologies for formal representation, interpretation, integration, and analyses of forest data. Approaches used in designing these ontologies are discussed in this section.

## 4.1. An Ontology Design Pattern for Spatial and Temporal Aggregate Data (STAD)

Ecological trends are easier to spot from aggregate data hence data aggregation is an integral part most ecology studies including forest ecology. For example, trends such as global warming are easier to spot from climate normals (30-year average of weather variables). Existing ontologies, such as the OGC standard on Observation and Measurement (O&M) [7], the Ontology of Units of Measure (OM) [8], Quantities, Units, Dimensions and Types (QUDT) [9] and a recent formalization of quantity kinds, values and units of measures (FOUnt) [10] semantically represent quantity kinds and units and values of measurements and quantities. However, they are insufficient for distinguishing between aggregate quantities.
Other ontologies, such as the Statistical Methods Ontology (STATO) [11] and the Ontology for Biomedical Investigations (OBI) [12], provide concepts to indicate the statistical techniques applied, such as taking an average or median. However, no existing ontology provides the spatial and temporal characteristics of aggregated quantities.

This work aims to develop an Ontology Design Pattern ([13], [14]) for Spatial and Temporal Aggregate Data (STAD) to fill this gap. Four qualities that are essential for fully expressing the semantics of aggregate quantities are identified: spatial support (where), temporal support (when), base quantity (what), and transformation kind (how). To ensure compatibility with existing vocabularies, existing ontologies of quantities are examined. QUDT was chose because it models quantity kinds as instances of the class QuantityKind, which allows explicit specification and comparison of the quantity kinds of instances of quantities as needed for the forestry data. Other quantity ontologies, such as OM represent quantity kinds instead as subclasses of Quantities, which would complicate expressing queries that compare the quantity kind of instances of quantities.

## 4.2. Forest Ecology Ontology (FEO)

This work also aims to develop a Forest Ecology Ontology (FEO) as an integrated domain ontology of forest inventories and related ecological factors. The development of FEO is guided by a set of sample competency questions including:

- Under what conditions does red spruce thrive?
- What characteristics have locations where both red and white spruce occur?

- What other locations (where we don't have tree occurrence information) have similar environmental characteristics as red spruce environmental preference?
- What tree species favor the condition in location X?
- Why can white ash not grow in location Y?
- What abiotic factor is most limiting to the occurrence of white ash?

Existing application of ontology in forestry can be classified into two main categories: (1) knowledge organization and management, which involves using vocabularis such as AGROVOC [15] for managing forest-related knowledge and datasets, and (2) semantic reasoning, which employs ontologies such as the Cross Forest Ontologies (https://crossforest.eu/) to express and query forest-related data using machine-readable language. Despite the wide use of environmental data in forestry, there is currently no ontology that provides comprehensive semantics for forest inventories and multiple environmental factors. There exists the environmental ontology (ENVO) [16] that provides a way to capture the semantics of environmental concepts, but it is limited to annotating environmental samples based on three criteria: biomes, environmental features, and environmental materials. This is not semantically rich enough for expressing detailed habitat descriptions such as the following example:

{910mm≤mean annual precipitation≤1320mm, -7°C≤January mean temperature≤-1°C, soil order = Spodosols or Inceptisols or Histosols, 4.0 ≤soil pH≤5.5}.

## 5. Research Results to Date

Currently, the two ontologies have been designed; an Ontology Design Pattern for Spatial and Temporal Aggregate Data (STAD) and the Forest Ecology Ontology (FEO). STAD is a quantities ontologies for expressing the semantics of Spatio-Temporal Aggregate Quantities such as Summer Mean Temperature. STAD reuses existing vocabularies form ontologies such as GeoSPARQL [17], OWL time ontology [18] and Ontology of Biomedical Investigations [12] to comprehensively express the sematics of aggregate quantities. Initial work on STAD is published in [19].
FEO has two fundamental branches; (1) Environmental Factors Module that expresses the definitions and inter-connectivity of environmental variables and (2) Environmental Preference Ontology Pattern that specifically connects organism species to their environmental preferences.

## Acknowledgments

Thanks to the developers of ACM consolidated LaTeX styles The presented material is based in part upon work supported by the National Science Foundation under grant OIA-1920908 for the project "Leveraging Intelligent Informatics and Smart Data for Improved Understanding of Northern Forest Ecosystem Resiliency (INSPIRES)". Torsten Hahmann has also been supported by NSF under grant OIA-2033607.

## References


[1] C. M. Malmstrom, Ecologists study the interactions of organisms and their environment, Nature Education Knowledge 10 (2010).

[2] FAO, In Brief to The State of the World's Forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies, FAO, Rome, 2022. URL: https://doi.org/10.4060/cb9363en.

[3] L. Yating, W. Zhenzi, X. Xiao, F. Hui, T. Xiaojia, L. Jiang, Forest disturbances and the attribution derived from yearly landsat time series over 1990–2020 in the hengduan mountains region of southwest china, Forest Ecosystems 73 (2021). doi:10.1186/s40663-021-00352-6.

[4] G. B. Bonan, Forests and climate change: Forcings, feedbacks, and the climate benefits of forests, Science 320 (2008) 1444–1449. URL: https://www.science.org/doi/abs/10.1126/science.1155121. doi:10.1126/science.1155121. arXiv:https://www.science.org/doi/pdf/10.1126/science.1155121.

[5] Y. Pan, R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, A. D. McGuire, S. Piao, A. Rautiainen, S. Sitch, D. Hayes, A large and persistent carbon sink in the world's forests, Science 333 (2011) 988–993. URL: https://www.science.org/doi/abs/10.1126/science.1201609. doi:10.1126/science.1201609. arXiv:https://www.science.org/doi/pdf/10.1126/science.1201609.

[6] M. G. Betts, C. Wolf, W. J.Ripple, B. Phalan, K. A. Millers, A. Duarte, S. H. M. utchart Taal Levi, Global forest loss disproportionately erodes biodiversity in intact landscapes, Nature 547 (2017). URL: https://doi.org/10.1038/nature23285. doi:10.1038/nature23285.

[7] S. Cox, ISO 19156:2011 - Geographic information – Observations and measurements, 2011. doi:10.13140/2.1.1142.3042.

[8] R. Hajo, A. Mark, T. Jan, Ontology of units of measure and related concepts, Semantic Web 4 (2013) 3–13. doi:10.3233/SW-2012-0069.

[9] R. Hodgson, P. J. Keller, J. Hodges, J. Spivak, Qudt - quantities, units, dimensions and types ontologies, 2014. URL: https://www.qudt.org/.

[10] B. Aameri, C. Chui, M. Grüninger, T. Hahmann, Y. Ru, The FOUnt ontologies for quantities, units, and the physical world, Appl. Ontology 15 (2020) 313–359. URL: https://doi.org/10.3233/AO-200231. doi:10.3233/AO-200231.

[11] A. Gonzalez-Beltran, P. Rocca-Serra, O. Burke, S.-A. Sansone, statistics ontology (stato), 2012. URL: http://stato-ontology.org/, accessed: July 21, 2022.

[12] A. Bandrowski, R. Brinkman, M. Brochhausen, M. H. Brush, B. Bug, M. C. Chibucos, K. Clancy, M. Courtot, D. Derom, M. Dumontier, et al., The ontology for biomedical investigations, PLoS ONE 11 (2016). doi:10.1371/journal.pone.0154556.

[13] A. Gangemi, V. Presutti, Ontology Design Patterns, 2009, pp. 221–243. doi:10.1007/978-3-540-92673-3_10.

[14] P. Hitzler, A. Gangemi, K. Janowicz, A. A. Krisnadhi, V. Presutti, Towards a simple but useful ontology design pattern representation language, in: WOP@ISWC, 2017. URL: https://api.semanticscholar.org/CorpusID:31642299.

[15] R. Lokers, Y. van Randen, R. Knapen, S. Gaubitzer, S. Zudin, S. Janssen, Improving access to big data in agriculture and forestry using semantic technologies, in: E. Garoufallou,

R. J. Hartley, P. Gaitanou (Eds.), Metadata and Semantics Research, Springer International Publishing, Cham, 2015, pp. 369–380.

[16] P. L. Buttigieg, E. Pafilis, S. E. Lewis, M. P. Schildhauer, R. L. Walls, C. J. Mungall, The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation, Journal of Biomedical Semantics 7 (2016) –. doi:10.1186/s13326-016-0097-6.

[17] Nicholas J. Car, Timo Homburg, Matthew Perry, John Herring, Frans Knibbe, Simon J.D. Cox, Joseph Abhayaratna, Mathias Bonduel, OGC GeoSPARQL - A Geographic Query Language for RDF Data, OGC Implementation Standard OGC 11-052r4, Open Geospatial Consortium, 2022. URL: http://www.opengis.net/doc/IS/geosparql/1.1.

[18] J. R. Hobbs, F. Pan, An ontology of time for the semantic web, ACM Transactions on Asian Language Information Processing 3 (2004) 66–85. doi:10.1145/1017068.1017073.

[19] K. Wiafe-Kwakye, T. Hahmann, K. Beard, An ontology design pattern for spatial and temporal aggregate data (stad), in: Proceedings of the 13th Workshop on Ontology Design and Patterns (WOP 2022) co-located with the 21st International Semantic Web Conference (ISWC 2022), CEUR- WS.org, online, 2022. URL: https://ceur-ws.org/Vol-3352/pattern4.pdf.