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

Kingsley Wiafe-Kwakye, Torsten Hahmann and Kate Beard

School of Computing and Information Science, University of Maine

Abstract

Many scientific disciplines heavily rely on statistically aggregated spatial and temporal data to describe, analyze and predict events and their interrelations. To help clarify and distinguish different kinds of statistical aggregations of temporally and spatially aggregated data, an ontology design pattern that aids in the specification of the semantics of such aggregations is presented. The ODP is specified in OWL and designed to guide the semantically correct fusing of spatio-temporally aggregated data and knowledge. Its use is illustrated using the climate normal of a Mean Summer Temperature.

1. Introduction

With advances in positioning techniques, sensor network technology, and remote sensing, spatio-temporal data about our environment has become increasingly available and opens up new opportunities for data analysis over larger geographic areas and multiple time spans. But the need to syntactically and semantically integrate data from multiple sources remains a major hurdle in realizing such large-scale research [15].

Scaling up environmental data analysis from local to regional or global levels heavily relies on aggregating data temporally and spatially. Generally, such data aggregation consists of applying statistical operations, such as average, minimum, maximum, sum, and count, to combine individual data points into summary statistics. It enables the processing of data in clusters rather than as individual data points and thereby reduces the amount of memory and processing power needed to further process such large scientific datasets. At the same time, aggregated data is easier to use for decision making. For example, trends such as global warming are easier to spot from annual summer temperature means than from daily or even hourly temperature readings. As an added benefit, data aggregation can also address privacy concerns by providing increased anonymity when compared to individual data points.

1.1. Motivation

While much progress has been made towards semantic interoperability of environmental data through the development of community-developed domain specific ontologies (e.g. EnvO [5]), the different ways of how spatial, temporal and spatio-temporal data are aggregated are still

largely unaddressed by existing ontologies. Two seemingly straightforward statistical measures with a rather precise-sounding label, such as "mean summer temperature" may be semantically incompatible because of differences in how the measures have been computed and what raw data has been used for it. More generic terms, such as "temperature", in ontologies or database row headers hide altogether whether the values are raw or aggregated measures and the kind of aggregations that have been applied.

Any aggregated measure, such as a mean summer temperature, is based on a set of other data – raw or aggregated. In this example, these underlying "base" data are typically a set of daily temperature values, which may themselves be statistical aggregations of more frequent measures. Daily temperatures, for example, are often the means computed using one of two approaches: twice-daily averaging (i.e., the average of the maximum and minimum temperature for that day) or hourly averaging (i.e., using the 24 hourly values of a day). Significant differences have been reported in the resulting daily mean temperatures [3], with some rapid weather events being linked to the daily mean temperature skewing towards the maximum or minimum daily temperature [3]. This distinction might affect a user's decision about which data to employ in a specific analysis, yet most data available on the web entirely lack the metadata that would convey to users how the aggregated data was calculated. If daily mean temperature is calculated via twice-daily averaging for certain states but hourly averaging is utilized for other states, an analysis that includes data from multiple states may be incorrect or biased as the result of the discrepancies in approach and granularity of the compared values.

1.2. Objectives

To help identify such issues, we aim to develop an Ontology Design Pattern (ODP) for spatial and temporal aggregate data (STAD) as a template for expressing the semantics of spatial, temporal, and spatio-temporal aggregated data more precisely. The ODP is intended to be used in annotating data and, later, comparing data from different datasets and spotting incompatibilities that may prevent data integration or require additional processing steps. The pattern development is motivated by concrete needs in the INSPIRES research project (https://crsf.umaine.edu/inspires/) to integrate different kinds of bioclimatic and soil-related forest data from New England into an integrated knowledge base – the "Digital Forest" – as a tool for improving our understanding of Northern Forest ecosystem resilience. The STAD ODP should specifically be able to represent the aggregations resulting from raster surface (such as satellite imagery) and raw point-based data. To construct the pattern, we first investigate what are central and distinct aspects of different kinds of spatial or temporal aggregations and thus must be semantically captured. The following questions guide the development:

- Is the data spatial and/or temporally aggregated?
- What statistical aggregation strategy has been applied to the data (average, minimum, percentile, standard deviation)?
- What is the spatial extent (i.e. different locations and their spatial distribution) of the aggregated data?
- What is the temporal extent (i.e. over what period of time and with raw data collected at what interval) of the aggregated data?
- What base quantities are used for the spatial and/or temporal aggregation?

2. Background and Related Work

A quantity is a measurable property of some object or collection of objects. Examples are the temperature in a room or the snow depth at a given location. Most scientific research concerns the measure and analyses of some quantities of some objects or events to discover trends and help understand both natural and artificial objects and occurrences. This requires understanding and communicating how quantities are measured, stored and shared.

Because various kinds of sensors are used to obtain measurements, ontologies of sensors and observations, such as the Semantic Sensor Network (SSN) ontology [7] and its revamped, more modular version [19] and the closely related Sensor, Observation, Sample, and Actuator (SOSA) ontology [16], can be used to describe, retrieve and share raw and statistical measures. A key concept in these ontologies is that of an observation, which encapsulates the idea of a sensor being used to measure or estimate the value of a property of an object or event. But this treats the sensor as a black box and its output as raw measurements, even if the sensor already performs some kind of aggregation. Comparison of whether data from different sensors can be integrated thus relies on comparing the devices, rather than the aggregation or computation. Our focus is on capturing the subsequent aggregations performed on sensor outputs.

But SSN, SOSA, as well as other ontologies that provide concepts for representing measures and quantities, such as the OGC standard on Observation and Measurement (O&M) [8], the Ontology of Units of Measure (OM) [12], Quantities, Units, Dimensions and Types (QUDT) [14] and a more recent formalization of quantity kinds, values and units of measures [1], all miss terminology to describe whether and how data is aggregated. Both OM and QUDT provide means to describe quantities with type, numerical value and unit of measurement, which may be enough for single quantities (i.e. raw data points) but are inadequate for describing distinctions between aggregate quantities.

While aggregating data before analysis or modeling is a common practice in many domains, there is a shortage of ontologies and patterns for describing spatially and temporally aggregated statistical quantities. The Statistical Methods Ontology (STATO) [11] is most closely related to STAD by providing a taxonomy of statistical methods (e.g. "arithmetic mean calculation") – which we reuse – and the aggregates produced by these methods (e.g. "average value"). STATO also provides a relation "computed_from" which links a computed quantity to its base quantity synonymous with our "hasBaseQuantity" relation. STATO, however, does not provide any spatial or temporal characteristics of aggregated quantities.

3. Conceptual Pattern

The Spatial and Temporal Aggregate Data (STAD) Ontology Design Pattern, of which we present a first iteration here, describes a unified framework for representing both individual and aggregate quantities. Aggregate quantities are described not only via the kind of transformation applied to the data but also by what base quantities are aggregated, and the critical temporal and/or spatial parameters that define *how* they are aggregated, as summarized in Figure 1.

At the highest level, we distinguish single quantity kinds, which represent raw measurements, from statistical quantity kinds, which represent quantities that are the outcome of applying
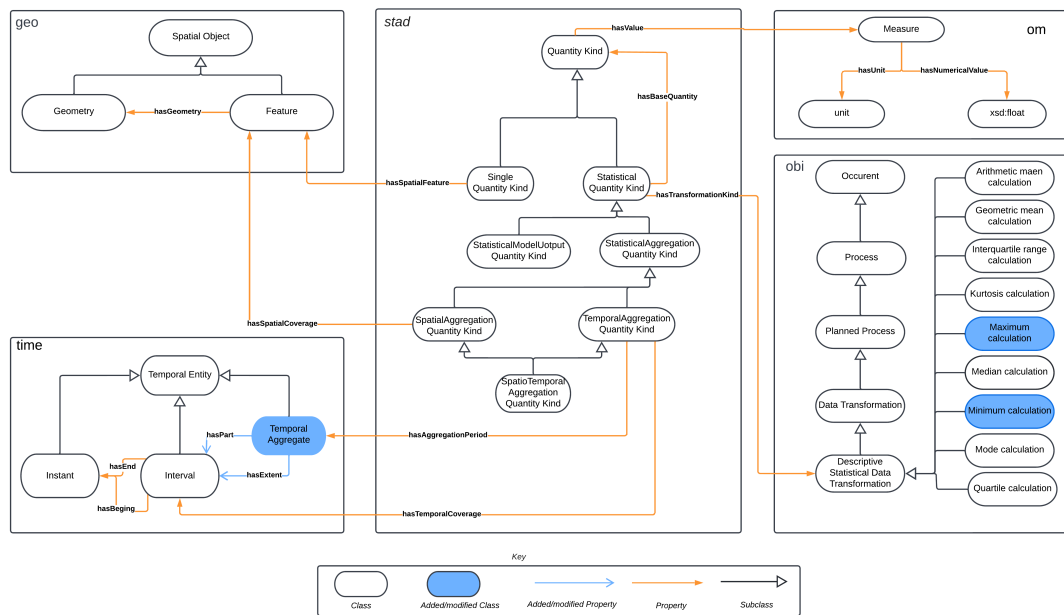
Figure 1: The essential characteristics of temporal, spatial and spatiotemporal aggregate quantities that STAD encodes are shown in the center, with connections to the existing ontologies Geosparql (geo)[17], OWL Time (time)[13], Observations and Measurements (om)[12], and Ontology of Biomedical Investigation (obi)[2] used to capture details of these characteristics.

some statistical transformation to a set of data points. Statistical quantity kinds are further categorized into: (1) model output quantity kinds (class *StatisticalModelOutputQuantityKind*) and (2) aggregate quantity kinds (class *StatisticalAggregationQuantityKind*). Model output quantities are produced by running some statistical model, e.g. a prediction model, with the base quantities as input. Aggregate quantity kinds capture the outcomes of simple statistical transformations that yield a summary statistic of the base quantities and are the focus of this paper. Examples include mean, mode, minimum, and variance of a dataset.

For environmental data, we're specifically interested in aggregated quantities that involve some spatial or temporal aggregation. In that respect, we distinguish three categories: data that are only spatially, only temporally, or both spatially and temporally aggregated (see the three subclasses of *aggregate quantity kinds*). Aggregation of a set of data points for a single resource (i.e. location) over multiple intervals of time or multiple time points forms a *temporal (time) aggregate*, while aggregation of data over a larger region or multiple locations but at a fixed time frame forms a *spatial (space) aggregate*. Finally, aggregating data over one or more regions or locations and a number of times produces a *spatiotemporal (space & time) aggregate*.

In the next section, we expand on the four characteristics that we have identified as essential for capturing the semantics of spatial temporal aggregates: spatial support (where), temporal support (when and, more precisely, how long and how frequently), base quantity (what), and transformation method (how). As seen in Figure 1, only spatiotemporal aggregates require all four characteristics, while only temporally aggregated data requires no spatial support and only spatially aggregated data requires no temporal support.

4. Ontology Formalization

In formalizing the pattern in this work, some relations and classes from existing ontologies have been reused (see Figure 1) to maximize interoperability of the STAD pattern with existing datasets and ontologies that use or expand those ontologies. In particular, the STAD pattern can be used in conjunction with existing ontologies for quantities and units, such as OM and QUDT, while offering concepts to explicitly capture the spatial and temporal characteristics of aggregate quantities for improved documentation of data provenance. To illustrate the use of the pattern, consider the encoding of a specific calculation of the summer mean temperature shown in Figure 2. We will use this example to present how the four key characteristics of spatial and temporal aggregate data are formally captured. The full formalization of the pattern and the example are provided at https://github.com/thahmann/spatialai/tree/master/Stad.
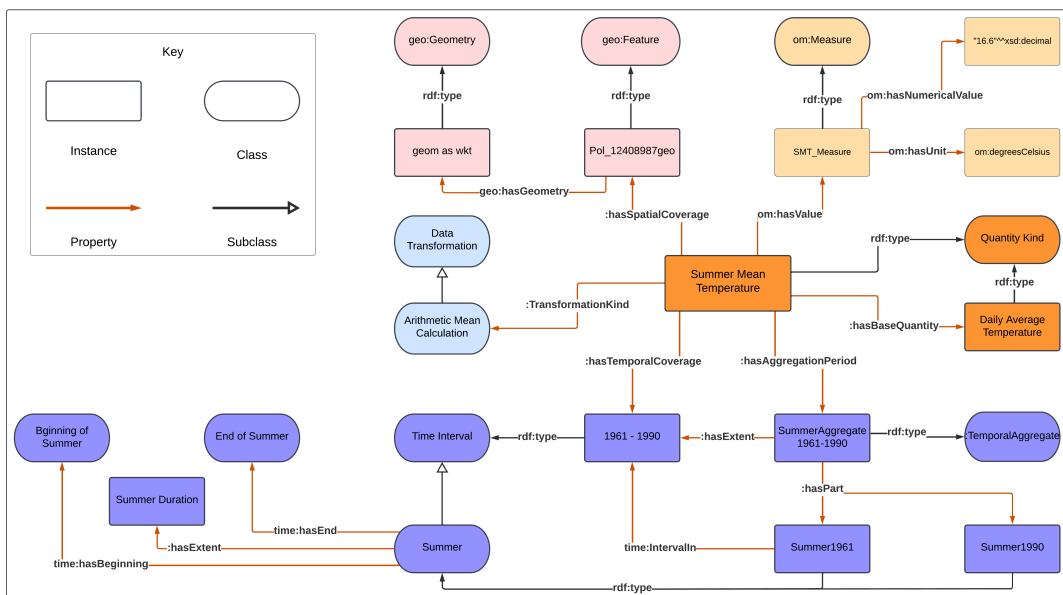
Figure 2: An instance of a summer mean temperature as a temporally aggregated quantity and its instance properties for the measure's unit and value (in yellow, upper right corner) and the four essential characteristics.

4.1. Temporal Aspects in Aggregations

To fully understand climate normals such as "summer mean temperature" or "winter mean temperature", one must understand two temporal variables. The first is the – often implicit – interval (e.g. 30 years) over which the normal has been calculated. In the example from Figure 2, "summer mean temperature" aggregates data from the years of 1961 to 1990. It is important to explicitly capture the aggregation interval because two summer means calculated during different years may actually differ if they used the immediate 30 years prior to their calculation. The second kind of temporal support concerns the time points (or subintervals) from which

data have been aggregated. For example, a summer mean temperature is calculated by using only the temperatures during a period defined as summer. Because the exact start and end of the interval referred to as "summer" could vary across organizations, regions or purposes (e.g. mereological summer vs. astronomical summer vs. summer growing season), it must also be made explicit. The two time variables are expressed by the relations *stad:hasTemporalCoverage* and *stad:hasAggregationPeriod*.

4.1.1. Aggregation Period

stad:hasAggregationPeriod is used to relate an aggregate to the temporal entities describing the included period of observation (e.g. every summer). The range of the relation is a subclass of *time:TemporalAggregate* the concept of a "Temporal aggregation" from the draft OWL Time extension ontology https://www.w3.org/TR/vocab-owl-time-agg/. In addition to the *time:hasPart* relation that relates an aggregate to a period therein, we capture the overall extent (which may be larger than the sum of parts) via the *stad:hasExtent* relationship:

stad:hasExtent rdf:type owl:ObjectProperty ;
  rdfs:domain stad:TemporalAggregate;
  rdfs:range time:ProperInterval.

stad:AggregationPeriod rdf:type owl:Class;
  rdfs:subClassOf time:TemporalAggregate,
    [rdf:type owl:Restriction;
    owl:onProperty stad:hasExtent;
    owl:cardinality "1"^^xsd:nonNegativeInteger;
    owl:allValuesFrom time:ProperInterval].

The property *stad:hasExtent* refers to a time period that is the complete extent of the aggregation whiles *stad:hasPart* connects to each sub-interval or time instant that is a component of the aggregation, in this case every summer from 1961 to 1990. The temporal intersection of the extent with all time intervals or instants that are part of the aggregation period produces the possibly non-convex time interval that precisely describes the aggregated time points (e.g. all summers that fall into the period 1961 to 1990 in the example from Figure 2). The relationship between the aggregated parts and the overall temporal extent is constrained by a property chain axiom on the *time:intervalIn* property from OWL Time [13] for time intervals and *stad:instantWithin*[1] for time instants.

time:intervalIn owl:propertyChainAxiom (stad:intervalPartOf stad:hasExtent).

time:instantWithin owl:propertyChainAxiom (stad:isExtentOf stad:hasInstantPart).

To be able to describe our example *stad:AggregationPeriod*, we need to first define what summer means in this particular dataset. Assuming the base quantities used for this aggregation

---

[1] *stad:instantWithin* is defined as the relationship between a time instant and a time interval such that the instant is either the beginning (*time:hasBeginning*) or end time (*time:hasEnd*) of the interval, or inside the interval (*time:inside*).

were collected from May 1 to September 30 of each year, we can define summer in OWL Time as follows:

ex:BeginningOfSummer rdf:type owl:Class;
  rdfs:subClassOf time:Instant;
  owl:equivalentClass [a owl:Restriction;
          owl:onProperty time:inDateTime;
          owl:allValuesFrom ex:BSDTD].

ex:BSDTD rdf:type time:DateTimeDescription;
  rdfs:subClassOf
        [a owl:Restriction;
        owl:onProperty time:day;
        owl:hasValue "—01"^^xsd:gDay],
        [a owl:Restriction;
        owl:onProperty time:month;
        owl:hasValue "–05"^^xsd:gMonth],
        [a owl:Restriction;
        owl:onProperty time:year;
        owl:someValuesFrom xsd:gYear].

Analogously, we can define the end of summer *ex:EndOfSummer.* To ensure that any specific summer covers only a single summer (and, for example, does not start in one year and ends after 17 months at the end of the next year's September), we explicitly enforce that any summer has a duration of 5 months.

ex:DurationOfSummer rdf:type time:Duration;
  time:numericDuration "5"^^xsd:decimal;
  time:unitType time:unitMonth.

We now use the definitions of the beginning, end, and duration of summer to define a class *Summer* as any time interval that begins on May 1 and ends on September 30 and has a duration of 5 months.

ex:Summer rdf:type owl:Class;
  rdfs:subClassOf time:ProperInterval,
  [owl:intersectionOf([a owl:Restriction;
         owl:onProperty time:hasBeginning;
         owl:allValuesFrom ex:BeginningOfSummer]
         [a owl:Restriction;
         owl:onProperty time:hasEnd;
         owl:allValuesFrom ex:EndOfSummer]
         [a owl:Restriction;
         owl:onProperty time:hasDuration;

owl:hasValue ex:DurationOfSummer]);
rdf:type owl:Class].

The summer aggregate can then be expressed by restricting its parts to only intervals that are summers:

ex:SummerAggregate1961-1990 rdf:type stad:AggregationPeriod,
[a owl:Restriction;
owl:onProperty time:hasInterval;
owl:allValuesFrom ex:Summer];
stad:hasExtent ex:yearsBetween1961-1990.
Its extent is defined in a straightforward way using OWL Time: *ex:yearsBetween1961-1990 a time:ProperInterval;*
time:hasBeginning ex:beginning1961;
time:hasEnd ex:ending1990.

ex:beginning1961 a time:Instant;
time:inXSDDateTimeStamp "1961-01-01T00:00:00-05:00"^^xsd:dateTimeStamp.

ex:ending1990 a time:Instant;
time:inXSDDateTimeStamp "1990-12-31T24:00:00-05:00"^^xsd:dateTimeStamp.

ex:pol_12408987tavesm_1961_90 stad:hasTemporalCoverage ex:yearsBetween1961-1990.

4.1.2. Temporal Coverage

stad:hasTemporalCoverage is then just the convex time interval that describes the overall temporal extent of the *TemporalAggregate*. For ease of use, we link a temporal (or spatio-temporal) aggregate directly to its temporal coverage via the *stad:hasTemporalCoverage* relationship, which is defined via a property chain axiom as follows:

stad:hasTemporalCoverage rdf:type owl:ObjectProperty ;
rdfs:domain stad:TemporalStatisticalQuantityKind;
rdfs:range time:ProperInterval;
owl:propertyChainAxiom (stad:hasAggregationPeriod stad:hasExtent).

4.2. Spatial Aspects in Aggregations

Spatial support is a term used in geostatistics to refer to the spatial unit used to sample the environment [18]. For spatial and spatiotemporal aggregates it is critical to track the spatial unit over which data are aggregated because its size and configuration can influence the distribution. The first-order statistics (such as central tendencies) of these distributions may be the same, but their second-order (such as variance) and higher-order statistics almost certainly would be different [9]. Also, it is critical to explicitly capture the spatial support to compare whether two

or more spatial aggregates (or base quantities) are about the same location and can be combined in a new aggregate.

STAD's relation *stad:hasSpatialCoverage* links an aggregated quantity to its spatial location that contains the locations of all its base quantities. We reuse GeoSPARQL's [17] *geo:SpatialObject* class with the subclasses *geo:Feature* and *geo:Geometry* to represent locations. *geo:Feature* is used to represent real world objects while *geo:Geometry* captures abstract geometric objects, such as points or ploygons, within a coordinate system. GeoSPARQL links features to their geometric representations via the *geo:hasGeometry* property. The spatial coverage of the example in Figure 2 can be expressed using GeoSPARQL as follows:

ex:pol_12408987 a geo:Feature;
  geo:hasGeometry ex:pol_12408987Geo.

ex:pol_12408987Geo a geo:Geometry;
  geo:asWKT "<http://www.opengis.net/def/crs/EPSG/8.5/4326> POLYGON
  *((-68.334 46.626,-68.335 46.6262,-68.335 46.627,-68.334 46.627,-68.337 46.626))"^^*
  <http://www.opengis.net/ont/geosparql#wktLiteral>.

ex:pol_12408987tavesm_1961_90 stad:hasSpatialCoverage ex:pol_12408987.

4.3. Base Quantity

To produce sensible aggregate quantities, only similar base quantities can be aggregated. STAD captures an aggregate's base quantities via the *stad:hasBaseQuantity* relation, where the range for specific aggregates can be restricted to any subclass of *quantity kind*, such as single quantity kinds or certain aggregate quantity kinds. For example, a mean temperature aggregate can be described to only aggregate base quantities that are temperatures themselves and have a similar spatial and/or temporal support.

4.4. Transformation (Aggregation) Kind

Several studies [4, 6, 10] have shown that the choice of aggregation technique that is applied to a dataset can impact the distribution and further analysis. Thus, it is critical to capture the kind of transformation applied for each aggregate quantity, which is accomplished by the *stad:transformationKind* relation that links an aggregate quantity to the aggregation technique. For describing different aggregation techniques, STAD reuses the *data transformation class* (OBI:0200000) provided by the Ontology for Biomedical Investigations (OBI) [2]. While OBI's data transformation class already defines several subclasses for some common statistical data transformation techniques, more can be added as needed. For example, we introduce *minimum calculation* and *maximum calculation*s as subclasses of OBI's *descriptive statistical calculation data transformation*.

5. Use Case

Most environmental data share common attributes. For instance, quantities could be of the same type, share spatial coverage, temporal coverage, aggregation period or transformation kind. For instance, several mean temperatures values would have one thing in common in order the meaningfully apply a *transformation kind* such as *arithmetic mean*. Annotating environmental data with relations specified by STAD allows us to define new classes as groupings of dataset by a common characteristic as illustrated in Figure 3. Grouping may be by location (e.g. *Mean Temperature_Bangor*, by temporal coverage (e.g. 1991 to 2020), or by aggregation period (e.g. all that use a definition of summer lasting from May 15 to September 30). Such defined subclasses help reduce redundancy in the annotations and ease information retrieval for answering questions such as the competence questions that guided the design of the STAD pattern.
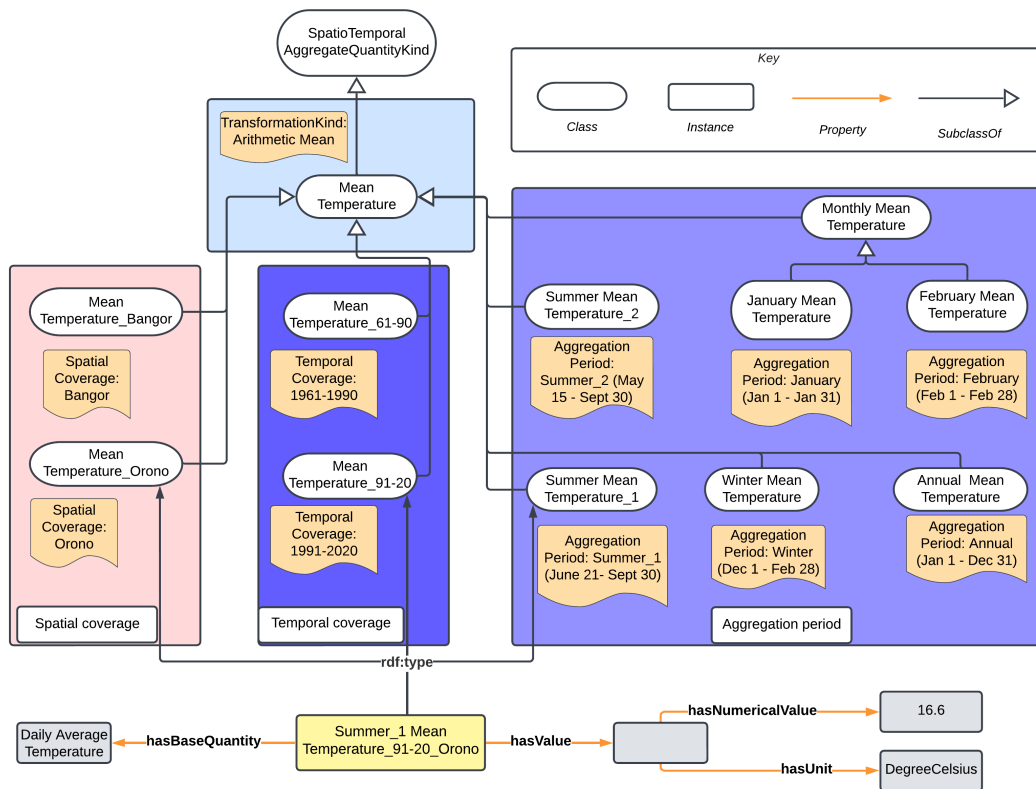
Figure 3: Using STAD to describe one specific kind of summer mean temperature (*Summer_1 Mean Temperature_91-20_Orono* shown at the bottom) as the arithmetic mean of daily mean temperature data for over the summers (June 21 to Sept. 30) from 1991 to 2020 at the location Orono, ME, which has a value of 16.6°C. An alternative definitions of summer as begins on May 15 is encoded as *Summer_Mean_Temperature_2*. The four colored boxed contain subclasses of *Mean Temperature* that fix one of the four aspects that describe a statistical aggregate.

6. Conclusion

This paper outlined STAD as an Ontology Design Pattern for spatially and temporally aggregated data. We identified key aspects for describing the semantics of a statistical aggregation and leveraged existing ontologies for time, space, statistical methods, and measurement quantities as much as possible. As next steps, we will test the ODP by annotating and disambiguating various kinds of statistical aggregations from bio-climatic variables in the INSPIRES project. We also plan to test the pattern's applicability to outputs of complex statistical models.

Acknowledgments

References

[1] Bahar Aameri, Carmen Chui, Michael Grüninger, Torsten Hahmann, and Yi Ru. The FOUnt ontologies for quantities, units, and the physical world. *Appl. Ontology*, 15(3):313–359, 2020. doi: 10.3233/AO-200231. URL https://doi.org/10.3233/AO-200231.

[2] Anita Bandrowski, Ryan Brinkman, Mathias Brochhausen, Matthew H. Brush, Bill Bug, Marcus C. Chibucos, Kevin Clancy, Mélanie Courtot, Dirk Derom, Michel Dumontier, et al. The ontology for biomedical investigations. *PLoS ONE*, 11(4), April 2016. doi: 10.1371/journal.pone.0154556.

[3] Jase Bernhardt, Andrew M. Carleton, and Chris LaMagna. A comparison of daily temperature-averaging methods: Spatial variability and recent change for the CONUS. *J. of Climate*, 31(3):979–996, February 2018. doi: 10.1175/JCLI-D-17-0089.1.

[4] Ling Blan and Rachael Butler. Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. *Photogrammetric Eng. & Remote Sensing*, 65(1):73–84, January 1991.

[5] Pier Luigi Buttigieg, Norman Morrison, Barry Smith, Christopher J Mungall, and Suzanna E Lewis. The environment ontology: contextualising biological and biomedical entities. *Journal of Biomedical Semantics*, 4(43):–, December 2013. doi: 10.1186/2041-1480-4-43.

[6] William A.V. Clark and Karen L. Avery. The effects of data aggregation in statistical analysis. *Geographical Analysis*, 8(4):428–438, 1976. doi: 10.1111/j.1538-4632.1976.tb00549.x.

[7] Michael Compton, Payam Barnaghib, Luis Bermudez, Raul Garcıa-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, et al. The SSN ontology of the W3C semantic sensor network incubator group. *J. of Web Semantics*, 17:25–32, 2012. doi: 10.1016/j.websem.2012.05.003.

[8] Simon Cox. Ontology for observations and sampling features, with alignments to existing models. *Semantic Web*, 8(3):453–470, 2016. doi: 10.3233/SW-160214.

[9] J. L. Dungan, J. N. Perry, M. R. T. Dale, P. Legendre, S. Citron-Pousty, M.-J. Fortin, A. Jako-mulska, M. Miriti, and M. S. Rosenberg. A balanced view of scale in spatial statistical analysis. *Ecography*, 25(5):626–640, August 2002. doi: 10.1034/j.1600-0587.2002.250510.x.

[10] Adam Errington, Jochen Einbeck, Jonathan Cumming, Ute Rössler, and David Endes-felder. The effect of data aggregation on dispersion estimates in count data models. *Int. J. Biostatistics*, 18(1):183–202, 2022. doi: 10.1515/ijb-2020-0079.

[11] Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, Orlaith Burke, and Susanna-Assunta Sansone. statistics ontology (stato), 2012. URL http://stato-ontology.org/. accessed: July 21, 2022.

[12] Rijgersberg Hajo, Assem Mark, and Top Jan. Ontology of units of measure and related concepts. *Semantic Web*, 4:3–13, January 2013. doi: 10.3233/SW-2012-0069.

[13] Jerry R. Hobbs and Feng Pan. An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing*, 3(1):66–85, mar 2004. doi: 10.1145/1017068.1017073.

[14] Ralph Hodgson, Paul J. Keller, Jack Hodges, and Jack Spivak. QUDT: Quantities, units, dimensions and types, 2011. URL https://qudt.org/.

[15] Nick J.B. Isaac, Marta A. Jarzyna, Petr Keil, Lea I. Dambly, Philipp H. Boersch-Supan, Ella Browning, Stephen N. Freeman, Nick Golding, Gurutzeta Guillera-Arroita, Peter A. Henrys, et al. Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, 35(1):56–67, October 2019. doi: 10.1016/j.tree.2019.08.006.

[16] Krzysztof Janowicz, Armin Haller, Simon J.D.Cox, Danh Le Phuoc, and Maxime Lefrançois. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *J. Web Semantics*, 56:1–10, May 2019. doi: 10.1016/j.websem.2018.06.003.

[17] Nicholas J. Car, Timo Homburg, Matthew Perry, John Herring, Frans Knibbe, Simon J.D. Cox, Joseph Abhayaratna, and 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] Richard E. Rossi, David J. Mulla, Andre G. Journel, and Eldon H. Franz. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, 62(2):277–314, June 1992. doi: 10.2307/2937096.

[19] Kerry Taylor, Armin Haller, Maxime Lefrancois, Simon Cox, Krzysztof Janowicz, Raul García-Castro, Danh Le-Phuoc, Joshua Lieberman, Rob Atkinson, and Claus Stadler. The semantic sensor network ontology, revamped. *18th International Semantic Web Conference, JT@ISWC 2019*, 2019.