Computers and Electronics in Agriculture 169 (2020) 105171

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A semantic approach for timeseries data fusion

Argyrios Samourkasidis, Ioannis N. Athanasiadis*

Wageningen University and Research, Droevendaalsesteeg 3, Wageningen 6708PB, the Netherlands

ARTICLE INFO

Keywords:
Environmental timeseries
Internet of Things
Legacy data
Semantic heterogeneity
Templates
FAIR data
Reasoning
Interoperability
Data reuse
APSIM
AgMIP
DSSAT
WOFOST

ABSTRACT

The data deluge following the rise of Internet of Things contributes towards the creation of non-reusable data silos. Especially in the environmental sciences domain, syntactic and semantic heterogeneity hinders data reusability as most times manual labour and domain expertise is required. Both the different syntaxes under which environmental timeseries are formatted and the implicit semantics which are used to describe them contribute to this end. Usually, the real meaning of data is obscured in a combination of short data labels, titles and various value codes, that require domain or institutional knowledge to decipher. The FAIR data principles for scientific data sharing are stewardship offer a framework based on community-adopted metadata. In this work, we present the Environmental Data Acquisition Module (EDAM) which focuses on data interoperability and reuse, and deals with syntactic and semantic heterogeneity using a template approach. Data curators draft templates to describe in an abstract fashion the syntax of the timeseries datasets they want to acquire or disseminate. They complement each template with a metadata file, which is used to annotate observables and their properties (including physical quantities and units of measurement) with terms from an ontology. EDAM employs a reasoner to infer compatibility among syntactically and semantically heterogeneous datasets, and enables timeseries, format and units of measurement transformation on-the-fly. Our approach utilizes a local ontology to store metadata about datasets, which enables EDAM to acquire and transform datasets which were originally stored with different semantics and syntaxes. We demonstrate EDAM in a case study where we transform meteorological input files of four agricultural models. Our approach, allows to cut across environmental data silos and facilitate timeseries reusability, as it enables users to (a) discover datasets in other formats, (b) transform them and (c) reuse them in their scientific workflows. This directly contributes to the toolshed for FAIR data management in environmental sciences. EDAM implementation has been released under an open-source license.

1. Introduction

Scientists and environmental practitioners nowadays are confronted with the vast array of legacy environmental datasets, that become available online, and also with new data produced via the Internet of Things (IoT) devices. Raw data must undergo certain modifications in order for new knowledge to be produced by environmental models (Rizzoli et al., 2007). However, transforming a dataset to be compatible with a certain data specification is a laborious process (Horsburgh et al., 2009) and usually requires a human expert intervention (Athanasiadis, 2015). This process hinders environmental data reusability (Ames et al., 2012), facilitates the formation of data silos (Terrizzano et al., 2015) and ultimately widens the data-to-knowledge gap (Elag et al., 2017).

Semantic heterogeneity among the legacy datasets hinders automatic data transformation. The interdisciplinary nature of environmental sciences impedes reusability which is essential in the era of (big) data (Rizzoli et al., 2007; Wilkinson et al., 2016). Environmental timeseries are typically curated by several organizations and are annotated with implicit semantics (Beran and Piasecki, 2009). The real meaning of the data is obscured in a combination of short data labels or titles combined with institutional knowledge (de Vos et al., 2017). Such implicit semantics concern the physical quantity that was measured (i.e. temperature); the units which were used (i.e. Celsius degrees), and the physical phenomenon (entity or process) that it was measured on (i.e. atmosphere surface air). Often there is implicit knowledge about temporal and spatial references and the observation and measurement protocol. As an example, atmosphere surface air temperature is typically measured with thermometers placed in shelters positioned two meters above ground, according to the World Meteorological Organization (WMO) specifications. Typically unit selection differs among countries, regions and even among scientific disciplines and domains (Gkoutos et al., 2012). The different ways observables are quantified with respect to units of measurement add to semantic heterogeneity. This can lead to errors in data reuse and interpretation (Horsburgh

* Corresponding author.
E-mail address: ioannis@athanasiadis.info (I.N. Athanasiadis).

https://doi.org/10.1016/j.compag.2019.105171
Received 17 September 2019; Received in revised form 18 December 2019; Accepted 19 December 2019
0168-1699/ © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

et al., 2009) and renders data transformation to other formats a rather manual process, which eventually hinders data reuse beyond disciplinary silos.

The FAIR data principles have been introduced to formalize the requirements that scientific data must adhere to, in order to become findable, accessible, interoperable and reusable, by both humans and machines (Wilkinson et al., 2016). A key component of the FAIR data management principles is the adoption of high-quality metadata using community-adopted standards. Utilizing ontologies to support semantic interoperability is not a new concept in the environmental data domain (Bowers and Ludäscher, 2004; Madin et al., 2007; Madin et al., 2008; Rizzoli et al., 2008; Athanasiadis et al., 2011; Gruber and Olsen, 1994). An ontology represents the knowledge of a certain domain in a formalized manner through a set of statements (axioms) that define concepts and relationships between concepts (Villa et al., 2009). In the environmental domain, an ontology has been used to identify the physical processes, quantities, and their attributes (e.g. units of measurement) in a standardized manner (Yu and Liu, 2015). There have been several ontologies related to environmental sciences developed in the past decades, which received rather limited adoption (Athanasiadis, 2015). There is also a movement to facilitate data interoperability and reusability through the creation of new ontologies and dictionaries, which will be suitable for the Web (Rijgersberg et al., 2011; Compton et al., 2012). However, no clear winner exists among all these ontologies. One potential way to cope with ontology heterogeneity is through the use of intermediate steps such as vocabulary alignment or ontology mediation.

Ontology mediation refers to the process of describing different datasets through one ontology so that a common context is created and values can be reused (Regueiro et al., 2017). It is used to integrate diverse datasets, each of which is described by a different ontology, in order to become interoperable and reusable (Wilkinson et al., 2016; Shu et al., 2015). A semantic reasoner is a software agent which is essential component in the ontology mediation process.) By definition, it infers the implicit relations of an ontology (Mishra and Kumar, 2011), but can also support mediation among a number of them (Bröring et al., 2011).

This work focuses on the interoperable and reusable principles for scientific data management. We present a declarative approach to cope with semantic heterogeneity in order to automate environmental timeseries processing and transformation. For each data file, we use a template to describe its syntax and a metadata file to annotate the corresponding observables through a vocabulary. Then, a semantic reasoner parses the metadata files and resolves relationships across the different data files. Data stored in a specific format can be automatically transformed to another syntax, with the reasoner inferring compatibility among the corresponding observables.[1] Also, we incorporated a unit of measurement transformation module. We demonstrate this with the weather input files of four crop modelling solutions, namely APSIM (Holzworth et al., 2014), AgMIP (Rosenzweig et al., 2013), DSSAT (Jones et al., 2003), and WOFOST (Diepen et al., 1989) and the meteorological timeseries data provided by the Koninklijk Nederlands Meteorologisch Instituut (KNMI).

The rest of the paper is structured as follows: Section 2 reviews contemporary approaches towards environmental data transformation and gives the background of template frameworks. Section 3 presents the objectives along with abstract architectural design of our approach and overviews its implementation. Section 4 demonstrates the application of the semantic approach and the used datasets. Finally, Section 5 discusses our initial key findings, identifies future work and concludes the research.

[1] We use the term *observable* as a working term for the various properties, parameters or variables stored in environmental data files. The term is not confined solely to properties that are directly observed, and may include derived ones, i.e. those calculated or inferred from others.

2. Background and related work

The ultimate objective of ontology-driven approaches is the integration of semantically heterogeneous datasets (Villa et al., 2009), i.e. the creation of a consolidated view of datasets that were originally curated differently, and annotated with different ontologies. This may enable having a single endpoint to submit queries to these heterogeneous datasets (Beran and Piasecki, 2009), providing seamless, frictionless access. In the environmental domain this process is described with many concepts: the terms mediation (Regueiro et al., 2017), translation (Shu et al., 2015) and integration (Leinfelder et al., 2010; Beran and Piasecki, 2009) are synonyms and have been used interchangeably. In the environmental data science literature we discern three approaches towards semantic interoperability, which are based on either:

1. approaches which support syntactic interoperability, e.g. environmental data management frameworks such as the ones offered by Open Geospatial Consortium (OGC) (Bröring et al., 2011) and CUAHSI (Ames et al., 2012), or spreadsheets (de Vos et al., 2017),
2. Semantic Web stack technologies (e.g. RDF datastores, SPARQL, etc.) (Ziébelin et al., 2017), and
3. scripts[2] that create custom-to-dataset solutions.

Usually, the last two approaches cope with both syntactic and semantic heterogeneity at once.

Transformation of syntactically heterogeneous environmental timeseries into a consistent format is the concept around environmental data management frameworks. These frameworks, such the OGC SOS (Bröring et al., 2012) and the CUAHSI HIS (Ames et al., 2012), cope with *syntactic heterogeneity* by hiding the implicit syntaxes of diverse datasets and offering them through consistent data models (e.g. O&M Cox, 2011, WaterML Taylor, 2014, etc.). Efforts have also been made towards supporting *semantic* interoperability of such well-established frameworks. Henson et al. designed a *semantic* extension for the OGC SOS in order to submit high-level queries to raw data (Henson et al., 2009). Reguiero et al. in Regueiro et al. (2017) use controlled vocabularies to align semantics of various data sources for semantic mediation tailored to the OGC SOS protocol. Beran and Piasecki developed a knowledge base on top of syntactically interoperable, CUAHSI WaterML formatted datasets. In Beran and Piasecki (2009), they related terms from local vocabularies which were used to annotate environmental datasets, with terms from a universal ontology. This way, they addressed semantic heterogeneity and provided an endpoint, called Hydroseek, to submit queries to heterogeneous datasets curated by various environmental agencies. While Hydroseek further ensures interoperability by exporting all data in a standardised MS Excel format, uses have to undergo additional transformations in order to use the data in their scientific workflows. Additionally, Hydroseek mainly supports data consumers, as the process to add new data repositories is not detailed and seems to be carried out by the platform creators.

The standardised structure offered by spreadsheets made their utilization popular in the environmental data science domain (de Vos, 2017). This structure accounts for syntactic interoperability, and thus efforts have been made in order to complement those with semantic capabilities. Shu et al. (2015) present an ontology-mediation approach to deal with the *translation* of environmental data encoded in spreadsheets into XML. De Vos et al. in de Vos et al. (2017) present their ontology mediation approach which concerns the annotation of natural spreadsheets using external vocabularies, in order to identify the domain model implicitly defined in these natural spreadsheets.

Semantic Web stack technologies, as Linked Open Data, allow for

[2] We define scripting as the process of creating custom (usually one-off) computer programs to deal with a specific task

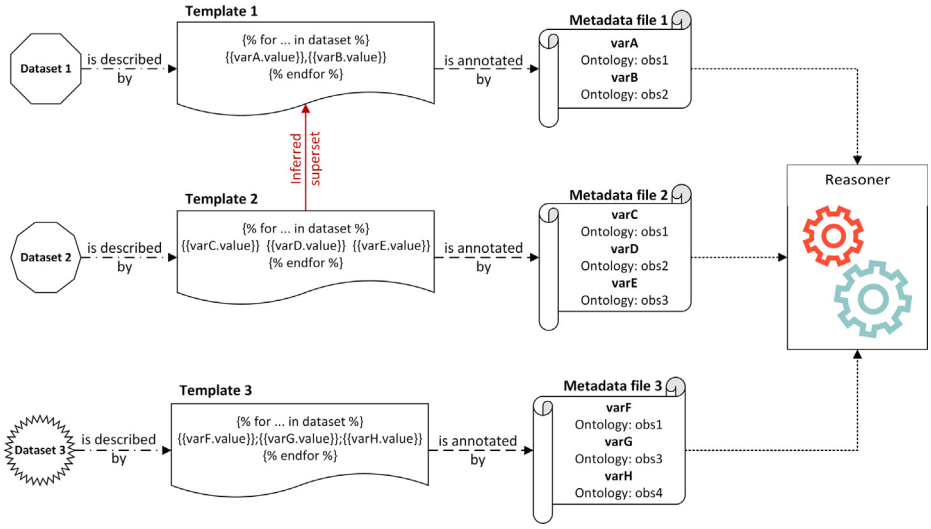


Fig. 1. The different shapes represent heterogeneous data syntaxes. For each different syntax, the template copes with *syntactic*, and the metadata file with *semantic* heterogeneity. The reasoner infers the compatibility of the data syntaxes based on the ontology definitions declared in the metadata files. The red line arrow depicts an inferred by the reasoner relationship between the templates of two different data syntaxes. *Dataset 2* can be *automatically* transformed to the syntax of the *Dataset 1*. It is an *extention* as it comprises of the same plus some extra observables. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

addressing both syntactic and semantic heterogeneity. The approaches which fall into this category, usually transcribe datasets into semantic-enabled datastores (triplestores) in order to support semantic data linking, processing and querying. For example Yu and Liu provide a single SPARQL endpoint to perform semantic queries to all underlying datasets (Yu and Liu, 2015). Bizer and Cyganiak present a tool, called D2R server, which publishes data stored in relational databases to a Semantic Web compatible format (Bizer and Cyganiak, 2006). Langegger et al. describe a mediator-based system for virtual data integration of scientific data (Langegger et al., 2008). Ziebelin et al. demonstrate a framework which uses the D2R server (Bizer and Cyganiak, 2006) to semantically link and integrate heterogeneous hydrological data sources. Interestingly, they support for enhanced interoperability as they disseminate the underlying, integrated datasets through OGC services (Ziébelin et al., 2017).

Environmental timeseries integration and transformation via *scripting* have been previously investigated within the agricultural domain. Porter et al. developed small software programs, called translators to transform the weather data files of four agricultural models into the AgMIP-consistent data format (Porter et al., 2014). Similarly, Woodard in Ag-Analytics developed Python scripts to acquire diverse datasets, store them into a consistent data schema and then offered the transformed data as a service (Woodard, 2016). In both works, the proposed solutions address the syntactic and semantic heterogeneity by aligning all datasets to a consistent data syntax with a common data model.

The work presented here is built-upon a mechanism which accounts for syntactic interoperability, and thus falls into the scope of the first approach. Associating a dataset with an abstract representation of its syntax contributes towards syntactic interoperability. Papoutsoglou et al. introduced the notion of using a template to describe a dataset syntax, parse the corresponding datapoints and offer them as services on the web (Papoutsoglou et al., 2015). In Samourkasidis et al. (2018) we designed and demonstrated a template framework for data acquisition to cope with *syntactic heterogeneity*. Using this framework, e-scientists without a strong computer science background can acquire and reuse environmental timeseries from various outlets (e.g. webpages, local files, databases, etc.) and create custom views of data using templates. In this work we extend this template framework with a declarative approach to cope with semantic heterogeneity. The benefits of our approach are described in the next section.

3. Methods

3.1. Objectives

There are three objectives in designing and developing a system to support automatic transformation of heterogeneous datasets. The first is to lower the environmental data science barriers, as the target users are e-scientists. As mentioned in Section 2, curating environmental datasets is a manual and custom process. In order to cope with semantic heterogeneity and interpret data, users should possess the implicit domain knowledge incorporated in environmental datasets. In this work, we embraced a declarative approach to cope with semantics, which does not require from users more technical skills than those they already have.

The second objective is to support the discovery of compatible datasets. We consider one dataset to be *compatible* with one other, only if the observables reported in the first are *equivalent* with those reported in the other. A semantic reasoner determines *compatibility*, based on semantic annotations provided by users. This enables users to find compatible datasets of interest, originally stored in other formats.

The third objective is *automatic* timeseries transformation between *compatible* formats. The automatic timeseries transformation to different formats consists of two steps: (a) syntax transformation, and (b) content transformation. The former concerns the layout transformation, such as the column order or naming. We focused on the latter, that is the unit of measurement transformation of the observables reported in a *source* dataset to match the ones of a *target* dataset. Our approach, allows to cut across environmental data silos and facilitate timeseries reusability, as it enables users to (a) discover datasets in other formats, (b) transform them and (c) reuse them in their scientific workflows. This directly contributes to the toolshed for FAIR data management in environmental sciences.

3.2. Abstract architectural design

There are three key-components involved in the design of our system: (a) template files, (b) metadata files and (c) reasoner. Fig. 1 depicts the interaction among the components. According to our approach, each distinct data syntax is represented through a template and a metadata file. The reasoner parses the metadata files, stores the ontology definitions for the reported observables in a local ontology and infers compatibility among their corresponding templates. The inferred relationships cope with the semantic heterogeneity, as they support for timeseries transformation among compatible syntaxes.

A *template file* is an abstract representation of a data file contents

using programming language agnostic semantics. Users draft one template for each data file *syntax*. They annotate important parts of the dataset using variables. Then, they define the observable metadata, represented through these variables in a metadata file.

A metadata file is bound to a single template and consists of semantic annotations for the reported observables. Users describe each observable through a name (e.g. Temperature), an ontology class (e.g. ontology:ObservableClass), and if applicable with qualifiers (e.g. max, min, daily). They also provide information about the corresponding units of measurement. For each unit of measurement, the name and symbol are mandatory fields, while a definition through an ontology class is optional. For both observables and units of measurement, users can define equivalent classes from other ontologies.

The reasoner parses the metadata files in order to infer *transformation compatibility* among the templates. Firstly, it creates an instance for each ontology class found in each metadata file. If applicable, it generates on-the-fly concrete subclasses to combine the abstract observable along with its related qualifier(s). For example, *maxDailyTemperature* is a *Temperature* subclass which combines a statistical (i.e. max) and a temporal (i.e. daily) qualifier. Secondly, it defines a new class for each template which is described by a general rule, called *axiom*. This axiom asserts in ontology language that the given template comprises of certain observables.

Next comes the unit of measurement transformation. The parser calculates the conversion factor between each set of the compatible *observables*. This calculation is based on the units of measurement which are defined in the source and target metadata files, accordingly. Finally, the conversion factors are applied on-the-fly (if applicable) on each column, and then transformed, according to the target template, in order for the dataset instance to be presented to the user.

3.3. Use of ontologies

We used a local ontology, which can map concepts and classes defined in different ontologies. This ontology comprises of three high level classes, *Observables*, *Qualifiers*, and *Templates*. In the *Observables* class, we create subclasses for the observables of each dataset, as defined by users in the metadata files. In this version, we annotated observables and units of measurement with classes from a custom, local ontology[3].

The *template variables* which are used to describe the dataset are stored as *instances* of the corresponding *Observable* subclass. We keep different namespaces for the *instances* of each template. This will enhance findability since each template will have its own prefix. So even for two templates using the same naming for their instances, there will be a distinction among them, based on the used prefixes (e.g. `AgMIP:rain` and `WOFOST:rain`). The namespaces can be optionally defined in the metadata file. In case they are missing, they can be generated based on the template file name.

The *Qualifiers* class is further refined into *Statistical* and *Temporal* mutually disjoint subclasses. Based on user input in metadata files, we define local statistical (e.g. max, min, mean, etc.) and temporal (e.g. daily, hourly, etc.) qualifiers and create their subclasses accordingly. A qualifier should always accompany an observable. In the end the system creates a composite subclass (e.g. *maxTemperature*) from the *Observable* (e.g. *temperature*) and its related *Qualifier(s)* (e.g. *max*) subclasses.

The *Templates* superclass holds the template definitions. We create a subclass for each distinct template along with its axiom definition. The axioms have direct reference to the *Observables* subclasses. The semantic reasoner uses these subclasses, when it comes to inferring compatibility among datasets.

Inferring compatibility among templates is facilitated by the local ontology and its properties. A *hasObservable* object property was defined to establish relationships among the *Templates* classes and their corresponding *Observables*. The axiom of a template with N associated observables defined with the *ontologyA*, is expressed in OWL language as follows:

```
Templates
and (hasObservable some ontologyA:observable1)
and (hasObservable some ontologyA:observable2)
...  and (hasObservable some ontologyA:observableN)
```

Based on the template *axioms* the reasoner infers four states of compatibility among two data syntaxes. If A is the *source* and B the *target* template representing different data syntaxes the possible states are:

a. A is **equal** to B, means that both templates comprise of the *same number* of *equivalent* observables.
b. A is an **extention** of B, means template A contains *all equivalent* observables reported in template B, plus one or more additional observables.
c. A is a **reduction** of B, is the reversed (b).
d. A is **non-compatible** to B, means that templates A and B may have or not observables in common.

A dataset represented with template A can *automatically* be transformed with template B in the first two cases.

3.4. Implementation

This approach extends the EDAM template framework Python module reported in Samourkasidis et al. (2018). It extends the template framework for data acquisition which already copes with syntactic heterogeneity, with a new module to support semantic operations. The system comprises of a parser and a semantic reasoner: EDAM supports the syntax transformation, Owlready2 Python library (Lamy, 2017) the ontology engineering and semantic reasoning, and Pint Python library (Grecco, 2019) the unit of measurement transformation. EDAM implementation is available via Python-pip under an open-source license.

We reused open source projects to provide further functionality. Specifically, we developed a parser to extract user definitions about observables and units of measurement from the metadata files, and utilized Owlready2 to store them in a local ontology. Additionally, Owlready2 supports the semantic reasoning to infer compatibility among the semantically heterogeneous datasets. We utilized Pint to support the unit of measurement transformation. Pint calculates the multiplicand factor of two units (i.e. source and target), based on their symbols. By design, Pint supports all SI symbols and their derivatives.

3.5. Limitations

The system presented here is intended for environmental timeseries. The system can handle the same file types as EDAM (Samourkasidis et al., 2018), i.e. text-based timeseries stored locally or remotely in one or more files, websites and/or relational databases.

Towards inferring compatibility among datasets, the system takes into consideration only the observable section in metadata files. The temporal (e.g. hourly, daily, etc) and/or statistical (e.g. min, max, mean, etc.) dimensions of the reported observables should be defined as *qualifiers*. By definition, observables that are reported in different temporal resolutions or regard different statistical value are not compatible. For example, the following sets of source to target transformations are (mutually) incompatible (a, b, c) and compatible (d):

a. *dailyTemperature* → *dailyMaxTemperature*,
b. *dailyTemperature* → *hourlyTemperature*,
c. *Temperature* → *dailyTemperature*,
d. *dailyTemperature* → *Temperature*,

[3] https://github.com/BigDataWUR/EDAM/blob/master/edam/resources/edam.owl

Table 1
The *implicit* semantics used by each data syntax to refer to the corresponding observables.

Observables	Datasets				
	APSIM	AgMIP	DSSAT	WOFOST	KNMI
Solar Radiation	radn (MJ/m^2)	SRAD (MJ/m^2)	SRAD (MJ/m^2)	irradiation (MJ/m^2)	Q (J/cm^2)
Avg temperature	–	–	T2M (°C)	–	TG (*d*°C)
Max temperature	maxt (°C)	TMAX (°C)	TMAX (°C)	maxt (°C)	TX (*d*°C)
Min Temperature	mint (°C)	TMIN (°C)	TMIN (°C)	mint (°C)	TN (*d*°C)
Precipitation	rain (mm)	RAIN (mm)	RAIN (mm)	precip (mm)	RH (dmm)
Wind speed	wind (m/s)	WIND (km/h)	WIND (m/s)	mwind (m/s)	FG (dm/s)
Relative Humidity	RH (%)	RHUM (%)	RH2M (%)	–	UG (%)
Dew Point Temperature	–	DEWP (°C)	TDEW (°C)	–	–
Vapor pressure	–	vprs (kPa)	–	emvp (hPa)	PG (dhPa)

```
!!!!   1/01/1961   to   31/12/2005
day   year   radn   maxt   mint   rain   wind   RH
273   2002   17.5   27.2   14.6   0   3.5   54
274   2002   13.6   23.1   14.7   0   5.3   40
275   2002   15.8   27.1   11.1   0   5.5   29
278   2002   15.2   23.1   15.2   0   3.4   47
```

(a) APSIM data file

```
!!!!   1/01/1961   to   31/12/2005
day   year   radn   maxt   mint   rain   wind   RH
{%for timestamp, radn,maxt,mint,rain,wind,RH in chunk%}
{{timestamp.dayofyear}} {{timestamp.year}} {{radn.value}} {{maxt.value}}
{{mint.value}} {{rain.value}} {{wind.value}} {{RH.value}}
{%endfor%}
```

(b) APSIM-specific template

```
Observables:
    - observable_id: mint
      name: Temperature
      ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#Temperature
      qualifiers: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#min
    - observable_id: maxt
      name: Max Temperature
      ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#Temperature
      qualifiers: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#max
    - observable_id: rain
      name: Rain
      ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#Rain
    - observable_id: radn
      name: Solar radiation
      ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#SolarRadiation
    - observable_id: wind
      name: Wind
      ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#WindSpeed
    - observable_id: RH
      name: Relative humidity
      ontology: https://github.com/BigDataWUR/EDAM/blob/features/semantics/semedam.owl#RelativeHumidity
Units of Measurement:
    - name: Millijoule per square meters
      symbol: mJ/m^2
      relevant_observables: radn
    - name: Percent
      symbol: \%
      relevant_observables: RH
    - name: Celsius
      symbol: degC
      relevant_observables: mint, maxt
    - name: Millimeters
      symbol: mm
      relevant_observables: rain
    - name: Meters per second
      symbol: m/s
      relevant_observables: wind
```

(c) APSIM-specific metadata file

Fig. 2. (a) An excerpt from a weather input file inspired by APSIM, (b) the corresponding *template* which represents the APSIM syntax and (c) the *metadata file* which annotates the variables used in (b) with concepts from a local ontology.

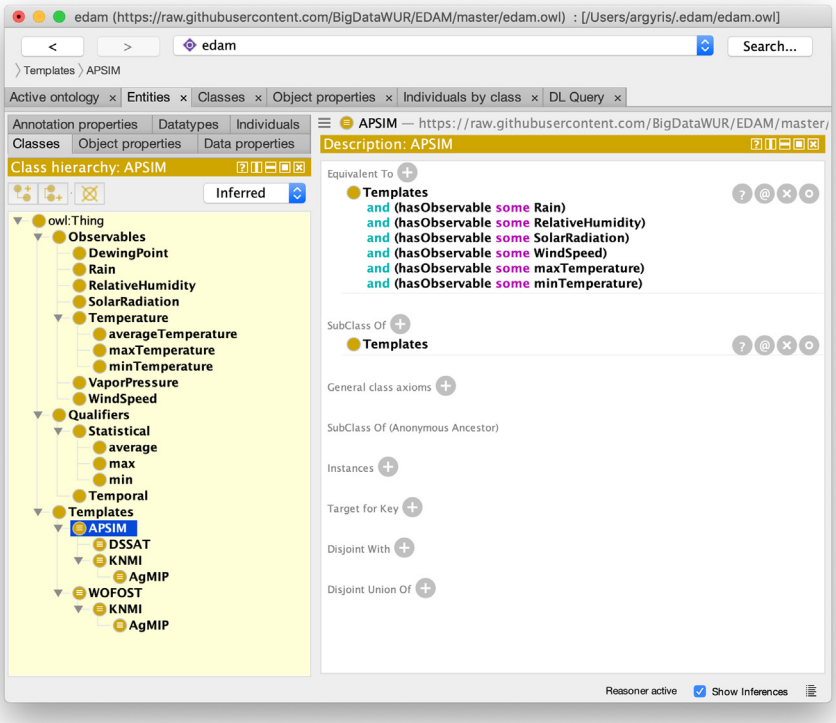


Fig. 3. A screenshot of the developed ontology in the Protege software. The *Observables* class consists of the observable types found in the different syntaxes. Combinations of these subclasses, describe each *Template* subclass. The reasoner inferred compatibility as depicted in the class-subclass hierarchy. For example, DSSAT dataset can be automatically transformed to APSIM, as the first is an extention (subclass) of the latter.

Users can refer to terms from external ontologies, but these are not directly imported. EDAM creates a local ontology with these terms which serves as a dictionary among the used terms. In this version, external ontologies are not imported to be further used.

The *automatic transformation* refers to the syntax and unit of measurement transformation. Any type of resampling in order to match source and target dataset temporal resolution is not included in the transformation process. Although EDAM offers this service, this is considered as a preprocessing step. Additionally, any possible spatial metadata are not considered when inferring compatibility.

4. Demonstration

4.1. Case studies

We demonstrate our semantic approach towards environmental timeseries transformation with the weather data files of four environmental models. Table 1 presents the selected datasets, the reported observables along with their implicit semantics and units of measurement. Besides the different semantics and units of measurement, each dataset has a different timeseries *syntax*.

For each dataset we developed a template to cope with the diverse *syntaxes* and a metadata file to annotate the reported observables. Fig. 2 depicts an excerpt of an input dataset for APSIM, the corresponding template (Fig. 2b), and the metadata file (Fig. 2c). The variable names inside the {{}} placeholders are used to draft the template, are reused in the metadata file to relate observables with their actual meanings. The observables are semantically annotated using a local ontology.

4.2. Compatible datasets

The reasoner operated on the five metadata files and updated the local ontology which can be further edited through dedicated ontology editors. It stored a class for each template, and automatically defined the template axiom based on the ontology classes of the related observables.

Fig. 3 is a screenshot of the Protege ontology editor (Gennari et al., 2003) which depicts the asserted and inferred relationships among the datasets. The class-subclass hierarchy depicts the relationships among the datasets. In principle, the subclass dataset is an *extention* of the parent class and thus automatic transformation is possible. Based on the template axioms, the reasoner inferred the following relationships:

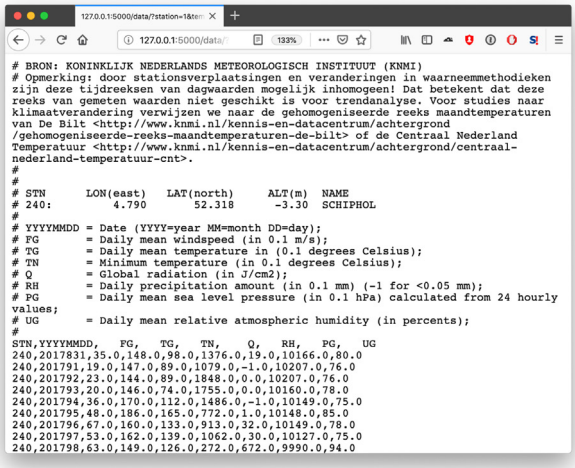
- DSSAT *extends* APSIM,
- AgMIP *extends* APSIM,
- KNMI *extends* APSIM,
- KNMI *extends* WOFOST,
- AgMIP *extends* WOFOST

Data compatibility was inferred based on the *combined* observable classes. These were generated on-the-fly: one subclass for every abstract observable. For example, for the AgMIP `TMAX`, the abstract observable is `Temperature` and the statistical qualifier is `max`. This combination results in the on-the-fly generation of `maxTemperature`, which is a `Temperature` subclass.

4.3. Automatic transformation

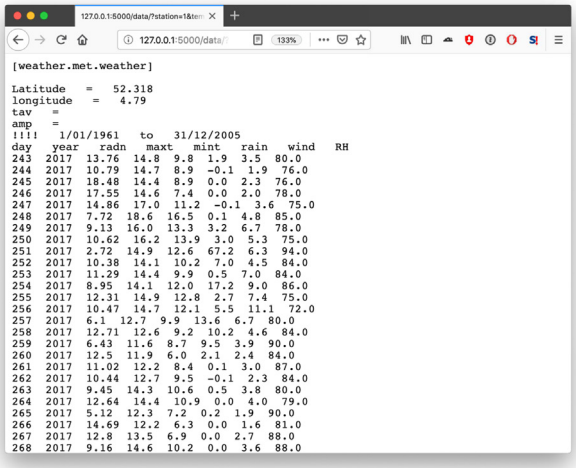
The system automatically transformed the compatible datasets upon user request. Transformation comprises of two parts: the syntax and semantic (or content) transformation. The former was performed by EDAM. The challenge here is with latter: the input and output templates use different semantics (i.e. observable identifiers). For example, AgMIP and APSIM datasets describe the max Temperature using the `TMAX` and `maxt` identifiers, respectively. The system established a relationship among the underlying observables of the input and output templates based on their compatibility. For example, it inferred that `maxt` and `TMAX` are synonyms and can be used interchangeably.

Unit transformation is performed on-the-fly upon dataset request. The system calculated the required conversion factors between source



```
# BRON: KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT (KNMI)
# Opmerking: door stationsverplaatsingen en veranderingen in waarneemmethodieken
zijn deze tijdreeksen van dagwaarden mogelijk inhomogeen! Dat betekent dat deze
reeks van gemeten waarden niet geschikt is voor trendanalyse. Voor studies naar
klimaatverandering verwijzen we naar de gehomogeniseerde reeks maandtemperaturen
van De Bilt <http://www.knmi.nl/kennis-en-datacentrum/achtergrond
/gehomogeniseerde-reeks-maandtemperaturen-de-bilt> of de Centraal Nederland
Temperatuur <http://www.knmi.nl/kennis-en-datacentrum/achtergrond/centraal-
nederland-temperatuur-cnt>.
#
#
# STN      LON(east)   LAT(north)     ALT(m)  NAME
# 240:         4.790       52.318      -3.30  SCHIPHOL
#
# YYYYMMDD = Date (YYYY=year MM=month DD=day);
# FG       = Daily mean windspeed (in 0.1 m/s);
# TG       = Daily mean temperature in (0.1 degrees Celsius);
# TN       = Minimum temperature (in 0.1 degrees Celsius);
# Q        = Global radiation (in J/cm2);
# RH       = Daily precipitation amount (in 0.1 mm) (-1 for <0.05 mm);
# PG       = Daily mean sea level pressure (in 0.1 hPa) calculated from 24 hourly
values;
# UG       = Daily mean relative atmospheric humidity (in percents);
#
STN,YYYYMMDD,   FG,   TG,   TN,    Q,   RH,   PG,   UG
240,2017831,35.0,148.0,98.0,1376.0,19.0,10166.0,80.0
240,201791,19.0,147.0,89.0,1079.0,-1.0,10207.0,76.0
240,201792,23.0,144.0,89.0,1848.0,0.0,10207.0,76.0
240,201793,20.0,146.0,74.0,1755.0,0.0,10160.0,78.0
240,201794,36.0,170.0,112.0,1486.0,-1.0,10149.0,75.0
240,201795,48.0,186.0,165.0,772.0,1.0,10148.0,85.0
240,201796,67.0,160.0,133.0,913.0,32.0,10149.0,78.0
240,201797,53.0,162.0,139.0,1062.0,30.0,10127.0,75.0
240,201798,63.0,149.0,126.0,272.0,672.0,9990.0,94.0
```

(a) KNMI formatted dataset



```
[weather.met.weather]

Latitude    =   52.318
longitude   =   4.79
tav    =
amp    =
!!!!   1/01/1961   to   31/12/2005
day   year   radn   maxt   mint   rain   wind   RH
243   2017   13.76   14.8   9.8   1.9   3.5   80.0
244   2017   10.79   14.7   8.9   -0.1   1.9   76.0
245   2017   18.48   14.4   8.9   0.0   2.3   76.0
246   2017   17.55   14.6   7.4   0.0   2.0   78.0
247   2017   14.86   17.0   11.2   -0.1   3.6   75.0
248   2017   7.72   18.6   16.5   0.1   4.8   85.0
249   2017   9.13   16.0   13.3   3.2   6.7   78.0
250   2017   10.62   16.2   13.9   3.0   5.3   75.0
251   2017   2.72   14.9   12.6   67.2   6.3   94.0
252   2017   10.38   14.1   10.2   7.0   4.5   84.0
253   2017   11.29   14.4   9.9   0.5   7.0   84.0
254   2017   8.95   14.1   12.0   17.2   9.0   86.0
255   2017   12.31   14.9   12.8   2.7   7.4   75.0
256   2017   10.47   14.7   12.1   5.5   11.1   72.0
257   2017   6.1   12.7   9.9   13.6   6.7   80.0
258   2017   12.71   12.6   9.2   10.2   4.6   84.0
259   2017   6.43   11.6   8.7   9.5   3.9   90.0
260   2017   12.5   11.9   6.0   2.1   2.4   84.0
261   2017   11.02   12.2   8.4   0.1   3.0   87.0
262   2017   10.44   12.7   9.5   -0.1   2.3   84.0
263   2017   9.45   14.3   10.6   0.5   3.8   80.0
264   2017   12.64   14.4   10.9   0.0   4.0   79.0
265   2017   5.12   12.3   7.2   0.2   1.9   90.0
266   2017   14.69   12.2   6.3   0.0   1.6   81.0
267   2017   12.8   13.5   6.9   0.0   2.7   88.0
268   2017   9.16   14.6   10.2   0.0   3.6   88.0
```

(b) KNMI dataset formatted according to APSIM

Fig. 4. The reasoner inferred compatibility between KNMI and APSIM and dataset depicted in (a) was automatically transformed according to APSIM format (b). The units of measurement transformation was performed on-the-fly.

and a target template units and applied them on the corresponding timeseries. Fig. 4 depicts a KNMI dataset (Fig. 4a) transformed according to the APSIM format (Fig. 4b). For this example, the conversion factors for the following unit of measurement transformation were calculated and applied on the source dataset:

$J/cm^2 \rightarrow MJ/m^2$

$d° C \rightarrow °C$

$dmm \rightarrow mm$

$dm/s \rightarrow m/s$

The system implementation is able to handle incompatible transformation requests, and annotations with unresolvable units of measurement. When a non-compatible transformation is attempted system issues an error. This error informs the user about the (in) compatibility of the involved datasets. The system can also handle units of measurement that either are not expressed correctly or are not SI units. In both cases, the system sets the conversion factor to 1 (i.e. no transformation) and raised warning messages to the user. For example, in this case a frequently found non-SI unit is the percent unit (%).

5. Discussion and conclusions

Environmental modelling solutions require their own input types and formats. As datasets are curated by different organizations, there are important differences in terms of *syntax* and *semantics*. Even related modelling solutions, such as APSIM and DSSAT (Jones et al., 2017), annotate the same observables through different local vocabularies and sometimes report their observables in different units of measurement. *Semantic heterogeneity* hinders environmental timeseries reusability, as transforming a dataset to another format is a laborious process (Beran and Piasecki, 2009; Horsburgh et al., 2009) which requires human expert intervention (Athanasiadis, 2015). In this work we presented a declarative approach to support environmental timeseries transformation. We employed a reasoner to infer *transformation compatibility* between semantically heterogeneous datasets, and developed a system to support units of measurement transformation. This contributes to the implementation of FAIR data management principles, and showcases the importance of metadata in automated discovery and transformation of data. We demonstrated how users may annotate datasets using a vocabulary, employ a reasoner, and transform data into other compatible formats.

Note that inconsistency when annotating observables with qualifiers leads to incompatibility. That is, qualifiers should be used either in all or in none of the datasets involved. While this statement is rather obvious, ensuring consistency when human users annotate manually the datasets is challenging. From the datasets we used only the KNMI included statistical and temporal (daily) qualifiers in their observable documentation. Had one independently annotated such datasets, the `KNMI:DailyMaxTemperature` would incompatible with `APSIM:maxT`, even though both refer to the very same observables. For this reason, we ignored the temporal and spatial metadata attributes for the KNMI data.

Automatic transformation of semantic heterogeneous data is essential towards environmental modelling in the IoT era.. This work also supports e-science, as the manual processing of data is often erroneous (Horsburgh et al., 2009). We also consider that this approach contributes towards lowering the e-science barriers (Swain et al., 2016). The proposed declarative approach copes with semantic heterogeneity, and enables e-scientists to transform compatible datasets to a given format, without developing scripts or being ontology engineers themselves.

This work has an exploratory character and sets the groundwork for future work. In this proof of concept, we supported semantic mediation, by enabling users to annotate the observables of the various datasets using a local ontology. The statistical and temporal qualifiers change fundamentally the meaning of the observables. However, most ontologies consider atomic measurements, thus statistical and temporal qualifiers are missing. Using qualifiers make observable annotation complete at a dataset level, and enables logical reasoning.

A possible direction for future work could be towards a module that infers temporal qualifiers and support automated annotation. Logically inconsistent annotations leads to incompatibility of datasets. Determining temporal qualifiers of an observable based on the contents of a data file is a step which would by-design offer consistent data representations, limiting human intervention.

Another possible direction for future work may be the design of *intermediate, semantic model-templates*. These model-templates would derive missing observables combining present ones. This is an essential step in cases where two syntaxes are not compatible because of a missing observable. An example from the case studies presented here is the incompatibility of the WOFOST and DSSAT datasets, because of the *Vapor Pressure* observable reported only in the former. An intermediate model-template to derive it, combining the *Temperature* and *Dew point* (present in the DSSAT) would allow automatic transformation between them.

5.1. Conclusions

In this work we presented our approach to cope with semantic heterogeneity towards transforming environmental timeseries data. We extended a data acquisition template framework which accounts for syntactic interoperability with a reasoner and a unit transformation module. This declarative approach enables users to annotate a data syntax once using terms from a vocabulary and then transform it to other compatible syntaxes. The employed reasoner infers the compatibility among different syntaxes, by creating a semantic description of each one. Then, the unit transformation module determines the relationship among the units and performs on-the-fly transformation where applicable. We demonstrated our declarative approach with the weather input files from four agricultural models and the meterological timeseries data from the Dutch Meteorological Office. In all cases where the reasoner inferred compatibility between two distinct datasets, we were able to transform the syntax and the content of one to another.

Declarations of Competing Interest

None.

Acknowledgements

This work was partially supported by the European Union's Horizon 2020 research and innovation programme grant 810775 (DRAGON). We are grateful to the two anonymous reviewers for their constructive feedback and comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compag.2019.105171.

References

Ames, D.P., Horsburgh, J.S., Cao, Y., Kadlec, J., Whiteaker, T., Valentine, D., 2012. HydroDesktop: web services-based software for hydrologic data discovery, download, visualization, and analysis. Environ. Model. Software 37, 146–156. https://doi.org/10.1016/j.envsoft.2012.03.013.

Athanasiadis, I.N., 2015. Challenges in modelling of environmental semantics. In: Environmental Software Systems. Infrastructures, Services and Applications. IFIP Adv. Inform. Commun. Technol. 448. Springer, pp. 19–25. https://doi.org/10.1007/978-3-319-15994-2_2.

Athanasiadis, I.N., Rizzoli, A.E., Donatelli, M., Carlini, L., 2011. Enriching environmental software model interfaces through ontology-based tools. Int. J. Adv. Syst. Stud. 4 (1/2), 94–105. https://doi.org/10.1504/IJASS.2011.042205.

Beran, B., Piasecki, M., 2009. Engineering New Paths to Water Data. Comput. Geosci. 354, 753–760. https://doi.org/10.1016/j.cageo.2008.02.017.

Bizer, C., Cyganiak, R., 2006. D2R server-publishing relational databases on the semantic web. In: Poster at the 5th international semantic web conference, Vol. 175, 2006.

Bowers, S., Ludäscher, B., 2004. An ontology-driven framework for data transformation in scientific workflows. In: Int. Workshop Data Integr. Life Sci. Springer, pp. 1–16. https://doi.org/10.1007/978-3-540-24745-6_1.

Bröring, A., Maué, P., Janowicz, K., Nst, D., Malewski, C., 2011. Semantically-enabled sensor plug & play for the Sensor Web. Sensors 11 (8), 7568–7605. https://doi.org/10.3390/s110807568.

Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., Lemmens, R., 2011. New generation sensor web enablement. Sensors 11 (3), 2652–2699. https://doi.org/10.3390/s110302652.

Bröring, A., Stasch, C., Echterhoff, J., 2012. OGC Sensor Observation Service 2.0, Implementation Standard 12–006. Open Geospatial Consortium.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Phuoc, D.L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K., 2012. The SSN ontology of the W3C semantic sensor network incubator group. J. Web Semant. 17, 25–32. https://doi.org/10.1016/j.websem.2012.05.003.

Cox, S., 2011. Observations and Measurements - XML implementation, Implementation Standard 10–025r1. Open Geospatial Consortium.

de Vos, M., 2017. Interpreting natural language spreadsheets. PhD thesis. Vrije Universiteit https://research.vu.nl/ws/portalfiles/portal/42791056/complete+dissertation.pdf.

de Vos, M., Wielemaker, J., Rijgersberg, H., Schreiber, G., Wielinga, B., Top, J., 2017. Combining information on structure and content to automatically annotate natural science spreadsheets. Int. J. Hum Comput Stud. 103, 63–76. https://doi.org/10.1016/j.ijhcs.2017.02.006.

Diepen, C., Wolf, J., Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. Soil Use Manag. 5 (1), 16–24. https://doi.org/10.1111/j.1475-2743.1989.tb00755.x.

Elag, M.M., Kumar, P., Marini, L., Myers, J.D., Hedstrom, M., Plale, B.A., 2017. Identification and characterization of information-networks in long-tail data collections. Environ. Model. Software 94, 100–111. https://doi.org/10.1016/j.envsoft.2017.03.032.

Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F., Tu, S.W., 2003. The evolution of Protege: an environment for knowledge-based systems development. Int. J. Hum Comput Stud. 58 (1), 89–123. https://doi.org/10.1016/S1071-5819(02)00127-1. http://www.sciencedirect.com/science/article/pii/S1071581902001271.

Gkoutos, G.V., Schofield, P.N., Hoehndorf, R., 2012. The units ontology: a tool for integrating units of measurement in science. Database. https://doi.org/10.1093/database/bas033. bas033.

Grecco, H.E., 2019. pint documentation, Readthedocs software documentation.

Gruber, T.R., Olsen, G.R., 1994. An Ontology for Engineering Mathematics. In: Doyle, J., Torasso, P., Sandewall, E. (Eds.), Principles of Knowledge Representation and Reasoning. Morgan Kaufmann, Bonn, Germany, pp. 258–269. https://doi.org/10.1016/B978-1-4832-1452-8.50120-2.

Henson, C.A., Pschorr, J.K., Sheth, A.P., Thirunarayan, K., 2009. SemSOS: Semantic sensor Observation Service. In: 2009 International Symposium on Collaborative Technologies and Systems. IEEE, pp. 44–53. https://doi.org/10.1109/CTS.2009.5067461.

Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., Moore, A.D., Brown, H., Whish, J.P., Verrall, S., Fainges, J., Bell, L.W., Peake, A.S., Poulton, P.L., Hochman, Z., Thorburn, P.J., Gaydon, D.S., Dalgliesh, N.P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F.Y., Wang, E., Hammer, G.L., Robertson, M.J., Dimes, J.P., Whitbread, A.M., Hunt, J., van Rees, H., McClelland, T., Carberry, P.S., Hargreaves, J.N., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., Keating, B.A., 2014. APSIM – evolution towards a new generation of agricultural systems simulation. Environ. Model. Software 62, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009.

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environ. Model. Software 24 (8), 879–888. https://doi.org/10.1016/j.envsoft.2009.01.002.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18, 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Brief history of agricultural systems modeling. Agric. Syst. 155, 240–254. https://doi.org/10.1016/j.agsy.2016.05.014. http://www.sciencedirect.com/science/article/pii/S0308521X16301585.

Lamy, J.-B., 2017. Owlready: ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif. Intell. Med. 80, 11–28. https://doi.org/10.1016/j.artmed.2017.07.002. http://www.sciencedirect.com/science/article/pii/S0933365717300271.

Langegger, A., Wöß, W., Blöchl, M., 2008. A semantic web middleware for virtual data integration on the web. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (Eds.), The Semantic Web: Research and Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 493–507. https://doi.org/10.1007/978-3-540-68234-9_37.

Leinfelder, B., Tao, J., Costa, D., Jones, M.B., Servilla, M., O'Brien, M., Burt, C., 2010. A metadata-driven approach to loading and querying heterogeneous scientific data. Ecol. Inform., vol. 5, 1, pp. 3–8, special Issue: Advances in environmental information management. https://doi.org/10.1016/j.ecoinf.2009.08.006. <http://www.sciencedirect.com/science/article/pii/S1574954109000685>.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., Villa, F., 2007. An ontology for describing and synthesizing ecological observation data. Ecol. Inform. 2 (3), 279–296. https://doi.org/10.1016/j.ecoinf.2007.05.004.

Madin, J.S., Bowers, S., Schildhauer, M.P., Jones, M.B., 2008. Advancing ecological research with ontologies. Trends Ecol. Evol. 23 (3), 159–168. https://doi.org/10.1016/j.tree.2007.11.007.

Mishra, R.B., Kumar, S., 2011. Semantic web reasoners and languages. Artif. Intell. Rev. 35 (4), 339–368. https://doi.org/10.1007/s10462-010-9197-3.

Papoutsoglou, E., Samourkasidis, A., Tsai, M.-Y., Davey, M., Ineichen, A., Eeftens, M., Athanasiadis, I.N., 2015. Towards an air pollution health study data management system-a case study from a smoky swiss railway. In: Johannsen, V.K., Jensen, S., Wohlgemuth, V., Preist, C., Eriksson, E. (Eds.), Adjunct Proc. 29th EnviroInfo and 3rd ICT4S Conference. University of Copenhagen, pp. 65–74 ISBN 978-87-7903-712-0.

Porter, C.H., Villalobos, C., Holzworth, D., Nelson, R., White, J.W., Athanasiadis, I.N., Janssen, S., Ripoche, D., Cufi, J., Raes, D., et al., 2014. Harmonization and translation of crop modeling data to ensure interoperability. Environ. Model. Software 62, 495–508. https://doi.org/10.1016/j.envsoft.2014.09.004.

Regueiro, M.A., Viqueira, J.R., Stasch, C., Taboada, J.A., 2017. Semantic mediation of observation datasets through Sensor Observation Services. Future Gener. Comput. Syst. 67, 47–56. https://doi.org/10.1016/j.future.2016.08.013. http://www.sciencedirect.com/science/article/pii/S0167739X16302722.

Rijgersberg, H., Wigham, M., Top, J.L., 2011. How semantics can improve engineering processes: A case of units of measure and quantities. Advanced Engineering Informatics 25, 276–287. https://doi.org/10.1016/j.aei.2010.07.008.

Rizzoli, A.E., Athanasiadis, I.N., Villa, F., 2007. Delivering environmental knowledge: a semantic approach. In: Proc. 21st International Conference on Informatics for Environmental Protection: EnviroInfo, pp. 43–50.

Rizzoli, A.E., Donatelli, M., Athanasiadis, I.N., Villa, F., Huber, D., 2008. Semantic links in integrated modelling frameworks. Math. Comput. Simul. 78 (2), 412–423. https://doi.org/10.1016/j.matcom.2008.01.017.

Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J., 2013. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. Agric. Forest Meteorol., vol. 170, pp. 166–182, agricultural prediction using climate model ensembles. https://doi.org/10.1016/j.agrformet.2012.09.011. <http://www.sciencedirect.com/science/article/pii/S0168192312002857>.

Samourkasidis, A., Papoutsoglou, E., Athanasiadis, I.N., 2018. A template framework for environmental timeseries data acquisition. Environ. Model. Software 117, 237–249. https://doi.org/10.1016/j.envsoft.2018.10.009.

Shu, Y., Ratcliffe, D., Compton, M., Squire, G., Taylor, K., 2015. A semantic approach to data translation: a case study of environmental observations data. Knowl.-Based Syst. 75, 104–123. https://doi.org/10.1016/j.knosys.2014.11.023.

Swain, N.R., Christensen, S.D., Snow, A.D., Dolder, H., Espinoza-Dvalos, G., Goharian, E., Jones, N.L., Nelson, E.J., Ames, D.P., Burian, S.J., 2016. A new open source platform for lowering the barrier for environmental web app development. Environ. Model. Software 85, 11–26. https://doi.org/10.1016/j.envsoft.2016.08.003.

Taylor, P., 2014. OGC WaterML 2.0: Part 1- Timeseries, Implementation Standard 10-126r4, Open Geospatial Consortium.

Terrizzano, I., Schwarz, P.M., Roth, M., Colino, J.E., 2015. Data wrangling: the challenging journey from the wild to the lake. In: Proceeding 7th Biennial Conference on Innovative Data Systems Research (CIDR), Online Proceedings, Asilomar, California, USA.

Villa, F., Athanasiadis, I.N., Rizzoli, A.E., 2009. Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. Environ. Model. Software 24 (5), 577–587. https://doi.org/10.1016/j.envsoft.2008.09.009.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, 'tP A C, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.

Woodard, J., 2016. Big data and Ag-Analytics: an open source, open data platform for agricultural & environmental finance, insurance, and risk. Agric. Finance Rev. 76 (1), 15–26. https://doi.org/10.1108/afr-03-2016-0018.

Yu, L., Liu, Y., 2015. Using linked data in a heterogeneous sensor web: challenges, experiments and lessons learned. Int. J. Digital Earth 8 (1), 17–37. https://doi.org/10.1080/17538947.2013.839007.

Ziébelin, D., Hobus, K., Genoud, P., Bouveret, S., 2017. Heterogeneous data integration using Web of Data technologies. In: Brosset, D., Claramunt, C., Li, X., Wang, T. (Eds.), Web and Wireless Geographical Information Systems. Springer International Publishing, pp. 35–47. https://doi.org/10.1007/978-3-319-55998-8_3.