Towards a Modular Ontology for Space Weather Research

Cogan Shimizu\textsuperscript{1}[0000–0003–4283–8701], Ryan McGranaghan\textsuperscript{2}[0000–0002–9605–0007], Aaron Eberhart\textsuperscript{1}[0000–0003–3007–5460], and Adam C. Kellerman\textsuperscript{3}[0000–0002–2315–936X]

\textsuperscript{1} Data Semantics Laboratory, Kansas State University, USA
\textsuperscript{2} Atmosphere and Space Technology Research Associates (ASTRA LLC), USA
\textsuperscript{3} Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, USA

Abstract. The interactions between the Sun, interplanetary space, near Earth space environment, the Earth’s surface, and the power grid are, perhaps unsurprisingly, very complicated. The study of such requires the collaboration between many different organizations spanning the public and private sectors. Thus, an important component of studying space weather is the integration and analysis of heterogeneous information. As such, we have developed a modular ontology to drive the core of the data integration and serve the needs of a highly interdisciplinary community. This paper presents our preliminary modular ontology, for space weather research, as well as demonstrate a method for adaptation to a particular use-case, through the use of existential rules and explicit typing.

1 Introduction

The solar-terrestrial system is complex, and consists of many interconnected subsystems nonlinearly interacting; energy flows from the surface of the Sun, through interplanetary space, into the near Earth space environment, and across the surface of the Earth. The expanse of the system as well as resource limitations dictate that a heterogeneous collection of data and models must be seamlessly utilized together [4].

Interactions across the solar-terrestrial system are manifest on Earth in myriad ways that pose threats to our technological infrastructure and ability to use the near-Earth space environment [7]. Among the most important, yet ironically, least well specified impacts is that of the electric power grid. During periods of enhanced space weather activity, a series of physical processes beginning with a solar event (e.g., the launch of a coronal mass ejection (CME) or a high speed stream (HSS) from the Sun) gives rise to intense electric currents reaching millions of Amperes surrounding the Earth, which then become electric currents on the ground flowing through electrical transmission lines. This phenomenon, known as Geomagnetically Induced Currents (GICs), can disrupt the operation of high-voltage power grid transformers via overheating and generation of harmonics, potentially leading to failures [6].
Power grid utility operators must stitch together observational and simulation data to know the appropriate responses to be taken. This is challenged by the number of disciplines that must cooperate and the intricate linking of the data across those disciplines involved – from space scientists, to earth scientists, to power system owners and operators, and power consumers.

What is needed is a semantic understanding of the notion linking solar events to power grid responses to form the foundation of an effective data integration system to study space weather and protect the power grid [5].

This paper describes the first modular ontological pattern that links the Sun to the power grid, specifically to guide decisions for power grid utilities.

2 Use-case Scenario

The Convergence Hub for the Exploration of Space Science (CHESS) project aims to unify the traditionally disparate communities and data sets that span the Sun, interplanetary space, near Earth space environment, the surface of the Earth, and the power grid. GICs arise from a series of interactions across these domains, beginning with the solar cloud of plasma interacting with the Earth’s magnetic field, creating currents in space and in the upper atmospheric region known as the ionosphere, which produces the electric field on the ground through magnetic induction. However, knowledge of many aspects of this chain is limited, largely due to the interdisciplinary nature of each interconnected link. Therefore, the motivating question is, “What is the ontological notion that links energy generated at the Sun, propagated through interplanetary space, and efficacious of disturbance to the power grid?” Important in this question is the responses that each link requires (e.g., conducting a predictive simulation, monitoring real-time data streams, or taking mitigative action to protect the grid) such that the ontological design must connect with ODPs for simulation activities, sensor networks, and the actions that must be taken by power grid utility operators. We detail the ontology designed for CHESS, demonstrating the value of the modular design approach to facilitate interoperability and reusability of existing ontological patterns.

3 The Preliminary Ontology

This section first introduces the Data Transformation ontology design pattern and its instantiation, the Simulation Activity Module, which are key pieces of the ontology. We then show how the module fits into the modular ontology for space weather. Next, we describe a mechanism for adapting the domain ontology to a particular use-case and, finally, provide a worked example.

The OWL files for the pattern, module, and ontology can be found online, as well as our supplementary data. The OWL files were generated using the
Towards a Modular Ontology for Space Weather Research

CoModIDE plugin [8] and annotated with OPLa [9] using the OPLa Annotator plugin [9]; both plugins are for Protégé [9].

3.1 The Data Transformation Pattern

In many fields, the transformation of data from one form to another, whether that be data-sanitation, analysis, or formatting, is a first-class operation. The Data Transformation pattern, herein described, allows for the description of data-driven workflows.

In January 2020, as part of a joint venture for the NSF’s Convergence Accelerator: Open Knowledge Networks initiative, several teams met to discuss the data modeling needs for each of the teams. It quickly became apparent that tracking data workflows, as well as availability of data, and even how that data might be used at all, was a very important aspect to each team’s use-case, as we see in Section 2. During this meeting we identified a number of interesting competency questions that serve to further frame the goals; a selection follows, with the entire document being found in the supplementary data.

CQ1. What datasets are available to view?
CQ2. What does dataset $X$ contain?
CQ3. In what ways is dataset $X$ used?
CQ4. What is the result of dataset $X$ transformed by Algorithm $A$?
CQ5. What dataset $X$ was used for input to Simulation $S$?

The resultant pattern can be shown in Figure 1. The core concept is Data Transformation. A DataTransformation always occurs in space and time (SpatiotemporalExtent) and in some ComputationalEnvironment. Each one also provides a number of roles: InputRole, OutputRole, and ParameterRole. These roles allow us to disconnect the Data from the transformation, acknowledging that the same data set can be used in different datasets and that no transformation is truly destructive. However, we do say that each instance of a role (as a transformation may have many, say, input data) may only be performed by exactly one Data. Further, every Data is an EntityWithProvenance, meaning that it is generally known from where data is sourced. We also specify, using the Explicit Typing pattern [10], that Data might have a DataType, which allows us to specify, in a domain dependent manner, broad categories of data or format of data.

We see in the schema diagram that Algorithm, ComputationalEnvironment, and EntityWithProvenance are represented by blue dashed boxes. This means that they are external to the pattern. That is, this pattern acknowledges that these concepts exist, and are integral to the notion of a DataTransformation, but makes no claims on their ontological structure. For the latter two, we recommend the use of their eponymous counterparts in [10]. Algorithm is very domain dependent and we leave it to be specified in specific use-cases.

---

5 See https://protege.stanford.edu/
6 The explicit declaration of types allows us to easily change domain dependent information, without impacting an ontology’s subsumption hierarchy. We posit that this strategy can aid with downstream alignment and co-reference resolution.
Axiomatization  This axiomatization was produced using the out-of-the-box axioms depicted in the Edge Inspector tool in CoModIDE [8], as well as the same axioms used during the systematic axiomatization process particular to Modular Ontology Engineering [11]. They are provided in description logic syntax below. The axioms are segmented to aid reading. Axioms 2-10 are scoped ranges (e.g. axiom 2) and existentials (e.g. axiom 3). Axiom 1 is a subclass. Axiom 11 is a structural tautology which serves to inform users of the intended use of an axiom. Axioms 12-15 and 29-32 are scoped domain axioms. Axioms 16-28 describe the roles that Data can perform.

Each of the properties included in this pattern have scoped domain and range, as we want to limit the ontological commitments made by this pattern. Each of the “performs” properties also have inverse existential and functionality. This means that only exactly one data can perform a role provided by a DataTransformation.

3.2 The Simulation Activity Module

The Data Transformation pattern is a necessarily general concept. During the ontology development process, we adapted the pattern to the use-case outlined in Sections 1 and 2. To do so, we made use of the strategy outlined in [1], which is called “Template-based Instantiation.” That is, we generally replaced names of classes and properties, instead of using subclass relationships. However, to
Towards a Modular Ontology for Space Weather Research

keep the provenance of the pattern’s structure, we use annotations from the Ontology Design Pattern Representation Language (OPLa) to indicate the original pattern. Figure shows the new module.

The major addition is concept of SimulatedData and additional specification of its own provenance. These changes to the pattern allow us to answer the following competency questions that are more specific to the use-case.

CQ6. What were the solar wind conditions at the time of a given GIC observation \( G \)?

CQ7. Is this GIC node at risk of exceeding a threshold GIC level in the next \( X \) time?

CQ8. What are the simulated conditions in given the results of a simulation?

The axioms for the additions are as follows.

\[
\exists \text{performsParameterRole.ParameterRole} \sqsubseteq \text{Data} \quad (29) \\
\exists \text{participatesInDataTransformation.DataTransformation} \sqsubseteq \text{Data} \quad (30) \\
\quad \exists \text{hasPayload.Payload} \sqsubseteq \text{Data} \quad (31) \\
\quad \exists \text{hasDataType.DataType} \sqsubseteq \text{Data} \quad (32)
\]
Fig. 1. The schema diagram for the DataTransformation pattern. Yellow boxes denote classes. Blue boxes with dashed outline denote external patterns or modules. Purple boxes denote controlled vocabularies. Solid arrows denote object or date properties, while open arrows denote the subclass relationship.

Fig. 2. The schema diagram for the SimulationActivity module. It utilizes the same notation as previous diagrams.

3.3 The Modular Ontology for Space Weather

At a top level, the purpose of the ontology is to tie together Activities, such as simulations or interpretations; Agents, such as power grid operators; Solar Events, such as coronal mass ejections; and Responses, such as simulate, monitor, or update. In addition, it is necessary to record the provenance of datasets as they are generated and transformed. We present the entire assembled ontology in Figure 3.

We have previously discussed the pattern used for one Activity: the Simulation Activity. The other activities are Monitor and Interpretation. The Monitor activity is currently left as a stub, which is a tiny pattern for acknowledging further complexity [10]. The Interpretation Activity is used when an Agent, either a person or computer algorithm, interprets some data to indicate that a Solar Event, or not, has occurred.

The Agent Role module is based on the Agent Role pattern, equipped with Explicit Typing for the categories of roles. This set of roles is a controlled vo-
Towards a Modular Ontology for Space Weather Research

Fig. 3. The schema diagram for the modular ontology for space weather research.

cabulary (CV) and would be particular to the domain. In the next section, we discuss how these CVs can be used to add structured domain knowledge.

Agents are responsible for certain Responses. In the same manner, there are different categories of Responses. Responses also encompass an activity. For example, a response to some solar event might be to simulate its impact, and the Activity it encompasses would be exactly that execution.

3.4 Adding Domain Knowledge with Existential Rules

Although the precise specification that an ontology provides is very helpful for modeling a complex scenario, it is also occasionally difficult to visualize exactly which connections are purely logical and which are essential for a user interacting with the ontology. In order to provide a more intuitive, representation we
show a First-Order Predicate Logic (FOPL) Rule statement that corresponds to entailments of some sample data that would comprise the ontology. As we are writing the rule in FOPL, we may omit many of the description logic classes that are conceptually important, but play a diminished role in the semantics for a domain expert, and instead focus on the connecting predicates that form the relevant entailment. In this example, we show how a SolarEvent, such as a “Solar Flare”, leads to a definite chain of Responses. Every solar event defined in the ontology must be represented in this way. This means that if another solar event is known by the ontology, a well-defined sequence of responses to that event must also exist.

The natural language for this example is: IF: Flare observed by Ramaty High Energy Solar Spectroscopic Imager (RHESSI) OR Solar Dynamics Observatory (SDO) THEN:

1. Simulate the radiation conditions at Earth using historical analog data (radiation flux at geostationary orbit for similar conditions)
2. Update radiation conditions using NOAA SWPC GOES satellites observations (real-time radiation flux at geostationary orbit)
3. Based on radiation conditions run ionospheric models (input data are simulated/real-time geomagnetic activity indices and high-latitude coupling parameters)
4. Based on ionospheric currents from ionospheric simulations, simulate geomagnetic field
5. Based on geomagnetic field from simulations, simulate geomagnetically induced currents (GICs)

\footnote{The abbreviations are as follows: NOAA is the National Oceanic and Atmospheric Administration, SPWC is the Space Weather Prediction Center, and GOES is Geostationary Operational Environment Satellites.}
hasSolarEventType(solarevent, “Solar Flare”) \to
\exists\text{sequentialresponse, response1, response2, response3, response4, response5, activity1, activity2, activity3, activity4, activity5}
requiresResponse(solarevent, sequentialresponse)\land
hasFirstResponse(sequentialresponse, response1)\land
\text{encompasses}(response1, activity1)\land
hasResponseType(response1, “Simulate radiation conditions at Earth using historical analog data”)\land
hasNextResponse(response1, response2) \land \text{encompasses}(response2, activity2)\land
hasResponseType(response2, “Update radiation conditions using NOAA, SWPC, and GOES satellites observations”)\land
hasNextResponse(response2, response3) \land \text{encompasses}(response3, activity3)\land
hasResponseType(response3, “Based on radiation conditions run ionospheric models”)\land
hasNextResponse(response3, response4) \land \text{encompasses}(response4, activity4)\land
hasResponseType(response4, “Based on ionospheric currents from ionospheric simulations, simulate geomagnetic field”)\land
hasNextResponse(response4, response5) \land \text{encompasses}(response5, activity5)\land
hasResponseType(response5, “Based on geomagnetic field from simulations, simulate geomagnetically induced currents”)

3.5 Example Triple Data

Below we reproduce a fragment of what the populated ontology would look like, corresponding to the rule in the previous section. Some triples are omitted for brevity and clarity, however no new data was created besides instances and strings used to fill out the ontology, so this can be viewed as a correct representation.

:Activity rdf:type owl:Class .
:Response rdf:type owl:Class .
:ResponseType rdf:type owl:Class .
:SequentialResponse rdf:type owl:Class .
:SolarEvent rdf:type owl:Class .
:SolarEventType rdf:type owl:Class .
:encompasses rdf:type owl:ObjectProperty .
:hasFirstResponse rdf:type owl:ObjectProperty .
:hasNextResponse rdf:type owl:ObjectProperty .
:hasResponseType rdf:type owl:ObjectProperty .
:hasSolarEventType rdf:type owl:ObjectProperty .
:requiresResponse rdf:type owl:ObjectProperty .
:asString rdf:type owl:DatatypeProperty .

:solarEvent rdf:type owl:NamedIndividual ,
:SolarEvent ;
:hasSolarEventType :solarEventType ;
:requiresResponse :sequentialResponse .
:solarEventType rdf:type owl:NamedIndividual ,
:SolarEventType ;
:asString "Solar Flare" .

:sequentialResponse rdf:type owl:NamedIndividual ,
:SequentialResponse ;
:hasFirstResponse :response1 .

:response1 rdf:type owl:NamedIndividual ,
:Response ;
:encompasses :activity1 ;
:hasNextResponse :response2 ;
:hasResponseType :responseType1 .
:response2 rdf:type owl:NamedIndividual ,
:Response ;
:encompasses :activity2 ;
:hasNextResponse :response3 ;
:hasResponseType :responseType2 .
:response3 rdf:type owl:NamedIndividual ,
:Response ;
:encompasses :activity3 ;
:hasNextResponse :response4 ;
:hasResponseType :responseType3 .
:response4 rdf:type owl:NamedIndividual ,
:Response ;
:encompasses :activity4 ;
:hasNextResponse :response5 ;
:hasResponseType :responseType4 .
:response5 rdf:type owl:NamedIndividual ,
:Response ;
:encompasses :activity5 ;
:hasResponseType :responseType5 .

:responseType1 rdf:type owl:NamedIndividual ,
:ResponseType ;
Towards a Modular Ontology for Space Weather Research

:asString "Simulate the radiation conditions at Earth using historical analog data" .

:responseType2 rdf:type owl:NamedIndividual ,
      :ResponseType ;
      :asString "Update radiation conditions using NOAA SWPC GOES satellites observations" .

:responseType3 rdf:type owl:NamedIndividual ,
      :ResponseType ;
      :asString "Based on radiation conditions run ionospheric models" .

:responseType4 rdf:type owl:NamedIndividual ,
      :ResponseType ;
      :asString "Based on ionospheric currents from ionospheric simulations, simulate geomagnetic field" .

:responseType5 rdf:type owl:NamedIndividual ,
      :ResponseType ;
      :asString "Based on geomagnetic field from simulations, simulate geomagnetically induced currents" .

:activity1 rdf:type owl:NamedIndividual ,
      :Activity .

:activity2 rdf:type owl:NamedIndividual ,
      :Activity .

:activity3 rdf:type owl:NamedIndividual ,
      :Activity .

:activity4 rdf:type owl:NamedIndividual ,
      :Activity .

:activity5 rdf:type owl:NamedIndividual ,
      :Activity .

4 Conclusion

This paper has introduced a modular ontology for supporting space weather research, especially in understanding the interactions between agencies and their responsibilities, as well as tracking the usage and provenance of data through observations and simulations. In doing so, we presented a novel Data Transformation pattern, which is further modified to fit our use-case via template-based instantiation to produce the SimulationActivity module. This is one of the core patterns that constitute the entire modular ontology. We then go to show how existential rules, making use of instance data in controlled vocabularies, can be used to add and frame domain knowledge. Finally, we discuss a worked example.

The next steps for this work will be the further development of the modules left as stubs, namely Observation Data, which has obvious connections to the

SSN/SOSA\(^7\) and the Monitoring activity. We will also be exploring deployment and integration of real world data.

Acknowledgements. The authors acknowledge partial support from the National Science Foundation under Grant Numbers 1936677 and 1937152, as well as thorough discussion with Sean Gordon, Lisa Kerr, and Philip Murphy.

References


See https://www.w3.org/TR/vocab-ssn/


