Ontology modeling with domain experts: The GeoVoCamp experience

Pascal Hitzler[1], Krzysztof Janowicz[2], and Adila A. Krisnadhi[1,3]

[1] Wright State University, OH, USA
[2] University of California, Santa Barbara, USA
[3] Faculty of Computer Science, Universitas Indonesia

Abstract. A series of GeoVoCamps, run at least twice a year in locations in the U.S., have focused on ontology design patterns as an approach to inform metadata and data models, and on applications in the GeoSciences. In this note, we will redraw the brief history of the series as well as rationales for the particular approach which was chosen, and report on the ongoing uptake of the approach.

1 Introduction: GeoVoCamps

Since Spring 2012, a series of GeoVocamps has been held in the U.S. which used a modeling approach driven by ontology design patterns (ODPs). In terms of topics they have mostly (but not exclusively) stayed close to the GeoSciences and related disciplines. Events have been held annually in Spring in Santa Barbara, CA, and in Fall in the eastern U.S., mostly in the Washington D.C. area, with additional occasional events in locations such as Notre Dame, IN, or Dayton, OH.[4] Each event usually drew between 20 and 30 participants, including 8-10 "regulars" which come very frequently to the events.

GeoVoCamps are *unconferences*, and as such are loosely structured, with the actual program decided upon by the participants on the spot. Starting with the Santa Barbara event 2012, the organizers suggested to use ontology design patterns as modeling approach, and indeed most of the work at these GeoVocamps has adopted this since.

2 Typical Workflow at a GeoVoCamp

Typically, organizers and some participants arrive at a GeoVoCamp with some ideas, and probably even some very crude sketches, on what could be modeled. Typically, these are geoscience notions of central importance, such as trajectory, spatiotemporal scope, or cartographic map scaling. The first half day is usually spent on introductory talks and presentation of topic suggestions. Participants then split into working groups of usually 8-10 people, to jointly develop a model for one of the suggested notions. Throughout the rest of the camp, thre are

[4] See `http://vocamp.org/wiki/Main_Page`.

frequent planary sessions in which reports on the ongoing work are presented and feedback is solicited from all participants. Sometimes some participants switch between groups based on personal interest and on the need to have a balance of experts in each of the groups.

Working groups ideally have a mix of participants: two or more domain experts, one or two participants familiar with base data addressed by the modeling, at least one ontology engineer with knowledge about formal axiomatizations and technical modeling choices, and at least one person who is familiar with the general ODP-based modeling process. Domain experts with differing perceptions of their topic are particularly stimulating for the discussions, as are data providers with different target use cases in mind: reconciling these differences within a single pattern usually leads to more versatile, and thus more readily reusable, patterns.

When a working group starts working on a model for a notion, then often the first half-day is used up by determining the scope of what shall be modeled, and by understanding and possibly reconciling different ideas which the group participants may have about the topic. This phase often feels almost unproductive, in particular for participants which did not have previous exposure to the process. After a few hours, however, the scope usually falls into place, and at this stage first rough drafts can emerge. The initial phase is often dominated by discussing the exact definition of terminology, by looking at concrete examples, in particular exceptional cases, by looking at real data and potential use cases, and sometimes by formulating concrete *competency questions* for a use case.

To give an example about this initial process: In some recent work regarding data from high-energy physics experiments at CERN's LHC [1], we originally set out to model the general notion of *final state* as in particle physics. After some time, though, it turned out to be more adequate and in tune with data and intended use cases to model a more specific notion of *detector final state*, which defines the physical characteristics that form the basis of measurements as presented in a published paper about findings from a particle physics experiment such as those located at the Large Hadron Collider (LHC) at CERN.

Most of the discussions during the remainder of a working group at a camp will then evolve around producing an informal graph which, for the ontology engineers, captures the main components of the notion in the sense that it informs the structure of the RDF graph which results from populating the pattern with data. To give an example of this type of informal graph, consider Figure 1 which is taken from [2]. It informs the RDF graph structure but it it still bare of details regarding RDF and OWL, e.g. namespaces, domain/range restrictions, and other axioms. In our experience, this format is ideal in order to engage with domain experts; of course the ontology engineers in the group will be aware of the missing details and will make sure that they gather enough details and knowledge to be able to fill in the more technical details after the event.

At the end of the 2-3 day event, a working group has then usually completed a stable draft of a pattern in the sense of having a stable graph, and having an understanding of how to populate the graph. The ontology engineers will also

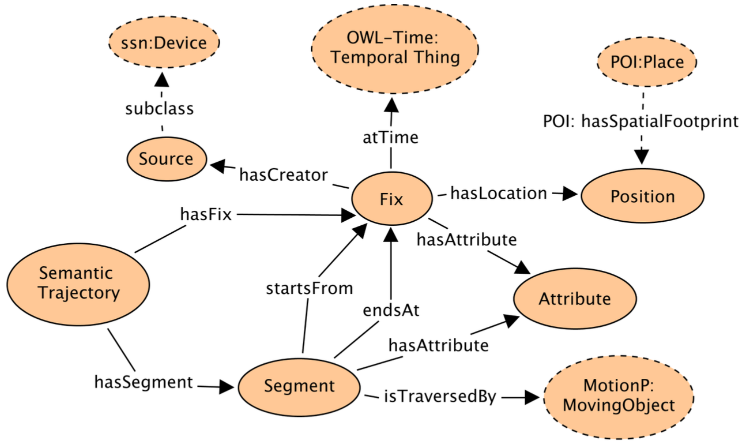

Fig. 1. Example graph for a pattern, taken from [2]

have made sure that they understand enough of the domain expert's perspectives to be able to come up with axiomatizations in OWL. A working group may then choose to work out further details offline, provide axiomatizations in OWL, as well as example populations with available data, and may publish the resulting documentation at a workshop, conference or journal. Sometimes work is also continued at a follow-up camp.

3 Spreading the GeoVoCamp Model

The interactive modeling approach taken at the GeoVoCamps as reported, has in the meantime been carried over into other events, communities, and disciplines mostly mediated by some camp participants. While we can in no way account for all the influences, we want to mention a few of the activities which were spawned.

Some of these actually resulted directly in camps which were targeted at specific application areas which were previously not considered. For example, for the 2015 Santa Barbara event[5] international experts on environmental impact of product life cycles – life cycle analysis or LCA for short – gathered to jointly look into metadata modeling for data integration and applications in their area. Two pattern descriptions already resulted from this, published at the 2015 Ontology and Semantic Web Patterns, WOP 2015, in Bethlehem, PA [3,6]. An event at University of Notre Dame held in summer 2015[6] drew experts from high-energy physics with involvement in CERN LHC experiments, as already mentioned above, and a corresponding pattern description has also already been published at WOP2015 [1].

[5] `http://vocamp.org/wiki/GeoVoCampSB2015`
[6] `http://vocamp.org/wiki/VoCampND2015`

The authors are furthermore all involved in the project *GeoLink*,[7] funded by the National Science Foundation under the EarthCube program, which has at its goal to showcase an integration infrastructure for earth science data repositories. The GeoVoCamp approach and ontology design patterns have been fully adopted by the ongoing project [4,5].

Frequent invited presentations to speak on the approach also indicate growing interest. E.g., such invited presentations have been made by different frequent GeoVoCamp participants for the ontolog forum,[8] at an OGC workshop,[9] at an RDA event,[10] for several EarthCube workshops and events,[11] at ESIP, the Federation of Earth Science Information Partners,[12] etc.

In our experience, one reason why the approach easily appeals to domain experts in application areas is that discussions are predominantly at the level of *knowledge transfer*: The domain experts are required to educate the ontology engineers about their most intricate terminoloty, and to do so to the minimum extent needed such that a formal model can be produced. At the same time, with suitable guidance by the ontology engineers, the domain experts can be shielded from technical discussions and from implementation details.

4 Bridging the Interdisciplinarity Gap

We want to briefly share some of our thoughts why our setting seems to be so effective for interdisciplinary modeling.

It is probably four things which come together in a very favorable way.

(1) The ODP approach prompts the group to work on one notion at a time. This provides a focus on a particular issue which enables the group to really explore the finer details of the notion. The informal graph which is collaboratively produced further emphasizes that the group is looking to develop precise and versatile definitions.

(2) Differences in perspectives by the domain experts in the group can be leveraged through their reconciliation into a single model. Hence they are particularly helpful in generating models which are versatile, in the sense that they capture all these different perspectives, and thus are general enough to be easily adopted or reused for other purposes or scenarios.

(3) The presence of ontology engineers in the group, which may have little or no knowledge about the subject matter which is to be modeled, can actually be turned to a significant advantage: In essence, the group needs to start by explaining to the ontology engineers what their notions really mean, and ontology engineers will ask for clarifications if the definitions given are not crisp enough to

[7] `http://www.geolink.org/`

[8] `http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2014_01_23`

[9] `http://earthcube.org/event/earthcube-session-ogc-quarterly-meeting-dc`

[10] E.g. `https://rd-alliance.org/rda-metadata-semantics-workshop-indianapolis-usa-feb-23-25-2015.html`

[11] E.g. `http://daselab.cs.wright.edu/pub/2014-05-DC-C4P.pdf`

[12] `http://daselab.cs.wright.edu/pub/2014-07-ESIP-ODP.pdf`

be developed into a formal model. This process will prompt domain experts to be precise, and to expose critical cases or exceptions to the general definitions, and it will also help to expose differences in perspectives by the domain experts. In our experience, it even sometimes exposes differences about which the participating domain experts were not aware earlier.

(4) Ideally, a working group also has participants which can act as bridges between the domain experts and the ontology engineers, i.e. participants who have some knowledge about either side. They can help explain when communication gaps occur, or if some notions have different meanings in different communities. Typical examples for the latter would be the meaning of “data” versus “metadata”, or the notion of “model.”

Throughout, it is paramount that participants understand their roles. Ontology engineers have to make sure they receive the information and minimal knowledge needed to produce the formal model after the meeting. Domain experts and data providers need to be aware that there are modeling experts in the room who have solid advice on how to structure a model. Bridge builders need to be aware that one of their main roles is that of facilitator and mediator. Listening to each other, and going with the flow of the group are very important in order to obtain useful results.

5 Tangible Outcomes

It is difficult to count all the outcomes – even the tangible ones – of an unconference. We easily came up with over a dozen papers which directly resulted from GeoVoCamp work, and almost a dozen which were directly influenced by it.[13] Of course these are primarily the papers where we were directly involved. And for many of the pattern papers the corresponding ontology design patterns have also been made available on the Web as OWL files.[14]

A significant body of work started in the working groups also awaits completion and casting into concrete patterns and papers. As these things go, without concrete deadlines or project demands, it depends solely on the priorities and bandwidths of the working group members whether this follow-up work will happen.

6 Conclusions

The GeoVocamps for ontology design pattern modeling have brought together an interdisciplinary community including ontology engineers and domain scientists from different disciplines, in order to engage in the modeling of conceptually strong and versatile ontology design patterns. We are seeing tangible outcomes, as well as a dissemination of the ideas and approach.

[13] `http://dase.cs.wright.edu/blog/geovocamps-taking-stock`

[14] E.g., on `http://ontologydesignpatterns.org/`.

We will continue the work, announcements can be found at `http://vocamp.org/wiki/Main_Page`, or contact us for inclusion in our distributor.

Acknowledgements. This work was supported by the National Science Foundation under award 1017225 *III: Small: TROn – Tractable Reasoning with Ontologies* and award 1440202 *EarthCube Building Blocks: Collaborative Proposal: GeoLink – Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences.*

References

1. Carral, D., Cheatham, M., Dallmeier-Tiessen, S., Herterich, P., Hildreth, M.D., Hitzler, P., Krisnadhi, A., Lassila-Perini, K., Sexton-Kennedy, E., Watts, G., Vardeman, C.: An ontology design pattern for particle physics analysis. In: Proceedings of the Workshop on Ontology and Semantic Web Patterns (6th edition), WOP2015, Bethlehem, PA, October 2015 (2015), to appear
2. Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., Hitzler, P., Dean, M., Kolas, D.: A geo-ontology design pattern for semantic trajectories. In: Tenbrink, T., Stell, J.G., Galton, A., Wood, Z. (eds.) Spatial Information Theory – 11th International Conference, COSIT 2013, Scarborough, UK, September 2-6, 2013. Proceedings. Lecture Notes in Computer Science, vol. 8116, pp. 438–456. Springer (2013)
3. Janowicz, K., Krisnadhi, A.A., Hu, Y., Suh, S., Weidema, B.P., Rivela, B., Tivander, J., Meyer, D.E., Berg-Cross, G., Hitzler, P., Ingwersen, W., Kuczenski, B., Vardeman, C., Ju, Y., Cheatham, M.: A minimal ontology pattern for life cycle assessment data. In: Proceedings of the Workshop on Ontology and Semantic Web Patterns (6th edition), WOP2015, Bethlehem, PA, October 2015 (2015), to appear
4. Krisnadhi, A.A., Hu, Y., Janowicz, K., Hitzler, P., Arko, R., Carbotte, S., Chandler, C., Cheatham, M., Fils, D., Finin, T., Ji, P., Jones, M., Karima, N., Mickle, A., Narock, T., O'Brien, M., Raymond, L., Shepherd, A., Schildhauer, M., Wiebe, P.: The GeoLink framework for pattern-based linked data integration. In: ISWC2015 Poster Proceedings (2015), to appear
5. Krisnadhi, A.A., Hu, Y., Janowicz, K., Hitzler, P., Arko, R., Carbotte, S., Chandler, C., Cheatham, M., Fils, D., Finin, T., Ji, P., Jones, M., Karima, N., Mickle, A., Narock, T., O'Brien, M., Raymond, L., Shepherd, A., Schildhauer, M., Wiebe, P.: The GeoLink modular oceanography ontology. In: Proceedings ISWC2015. Lecture Notes in Computer Science, Springer (2015), to appear
6. Yan, B., Hu, Y., Kuczenski, B., Janowicz, K., Ballatore, A., Krisnadhi, A., Hitzler, P., Suh, S., Ingwersen, W.: An ontology for specifying spatiotemporal scopes in life cycle assessment. In: Proceedings of the Diversity++ Workshop, Bethlehem, PA, October 2015 (2015), to appear