

Introduction: Ontology Design Patterns in a Nutshell

Krzysztof Janowicz, University of California, Santa Barbara, USA

Aldo Gangemi, LIPN, Université Paris 13, CNRS UMR7030 and ISTC-CNR, Italy

Pascal Hitzler, Data Semantics Laboratory, Wright State University, Dayton, OH, USA

Adila Krisnadhi, Data Semantics Laboratory, Wright State University, Dayton, OH, USA; and Faculty of Computer Science, Universitas Indonesia

Valentina Presutti, STLab, ISTC-CNR, Italy

Patterns in general can be defined as distinctive and repetitive invariants across observed data, objects, processes, and so forth, that are either manmade or occur naturally. Design patterns have emerged in computer science from the pioneering architectural work of Christopher Alexander [1], firstly applied to software engineering [9], then to workflows [31], HCI [30], data modeling [17], knowledge engineering [6], and eventually the Semantic Web [10, 13], where they are known as *Ontology Design Patterns* (ODP), *knowledge patterns*, or *linked data patterns*, depending on the community that uses them, e.g., ontology designers, knowledge engineers, linked data publishers, and so forth.

Patterns on the Semantic Web typically emerge from (linked) data, ontologies, and queries, as well as from procedural aspects of design at either the modeling or implementation stage. The main innovation of design patterns lies in the observation that a majority of datasets and ontologies share a relatively small set of common modeling and publishing challenges that can be approached by a common strategy, established best practice, or by combining existing building blocks. In principle, there is a distinction between design patterns as critical reviews of alternative ways of modeling something, such as a review of alternatives to represent an n -ary relation with $n \geq 2$ in RDF, vs. design patterns as reusable components to create a product, such as the description of a reusable ontology to represent such an n -ary relation in OWL. However, a proper design pattern should always mention its motivating requirements, applicability limits, benefits and shortcomings, and so forth. Therefore, whether it focuses on alternative methods, or on one or multiple alternative reusable components, a design pattern always provides a critical approach to deal with recurrent problems. Finally, it

is worth pointing out that the initial definition of patterns as invariants applies to any kind of pattern, while it makes sense to distinguish the purely symbolic patterns of mathematical pattern science [14], as studied in data mining, machine learning or complex systems, from knowledge patterns. Knowledge patterns are not merely symbolic, they also have a semantic interpretation, be it formal, or cognitive; e.g., a type of fact reported in the news.

While ontology design patterns can be considered analogous to software design patterns known from object oriented modeling, there are also clear differences. For instance, while ODPs and software design patterns can act as strategies, there is no clear design pattern counterpart for so-called *content ODPs* (knowledge patterns). These are typically modeled for frequently reoccurring aspects of more complex ontologies and thus act as building blocks rather than strategies. Examples include the *trajectory pattern* [18] that can be used in ontologies that model the transportation domain, wildlife monitoring, scientific cruises, to name but a few. Such building blocks, however, are not confined to domain-specific cases. For example, the *information realization pattern* [25] addresses the common problem of describing the relation between an information object, e.g., a book, and its physical realization, i.e., the specific paper copy of said book.

While the number of strategy patterns, such as the *structural n-ary relation pattern*, is relatively small, there is a wide and growing variety of knowledge patterns, and no immediate reason to establish an upper bound for their number. This leads to an interesting question, namely whether and how knowledge design patterns can be distinguished from other *small* ontologies. There are many possible criteria (cf. [10, 13, 25] for discussions). Ideally, ontology design patterns should be extendable but self-contained, minimize ontological commitments to foster reuse, address one or more explicit requirements (or use cases, competency questions), be associatable to an ontology unit test [33], be the representation of a core notion in a domain of expertise (so-called “blinking effect”, [13]), be grounded in conceptual or lexical frames [24], be alignable to other patterns, span more than one application area or domain, address a single invariant instead of targeting multiple reoccurring issues at the same time, follow established modelling best practices, and so forth.

While these criteria imply that there is a smooth transition between small ontologies and patterns, they also highlight the fact that ontology design pattern research itself is related to numerous other areas of study such as ontology modularization [8], ontology alignment [28], Linked Data publishing [27], knowledge extraction [24], knowledge discovery [22], ontology design [3], foundational ontologies, and so on. In many of the aforementioned cases, OPDs have provided new perspectives and insights to fuel ongoing research. In addition, patterns have also served as minimal core components of widely used ontologies such as the Semantic Sensor Network ontology [7]. Finally, some influential ontologies, such as the Simple Event Model [32] have either inspired modern ODP research, or could be perceived as patterns themselves.

This leads to another interesting question, namely why we are recently witnessing a rapid increase in the publication and reuse of patterns more than 10 years after their initial excogitation. We believe that this is due to a combination of several factors, more specifically changes in the Semantic Web research land-

scape. Early work on the Semantic Web (ca. 2001-2008) was heavily driven by foundational investigations of information ontologies and knowledge representation languages. Consequently, the focus was rather on describing which ontological choices can and should be made in partitioning the world, leading to questions of how to distinguish objects from processes, designing well modularized core ontologies in, say, the medical or legal domains [11, 12], or identifying *anti-patterns* for educational and model-checking purposes [5, 23, 26]. In many respects, the early Linked Data work (ca. 2008-2012) can be seen as a direct counter-reaction to address those problems by proposing minimal vocabularies and heavily restricting the expressivity of used knowledge representation languages, to focus on scalability. While this led to a rapidly growing ecosystem of interconnected data hubs, the early Linked Data Web could not fulfil some of the key promises it was set out to address – federated queries across distributed query endpoints being the most prominent of them. In a nutshell, *linked data* without semantics turned out to be merely *more data* [2, 13, 19], and required a detailed manual inspection and familiarity with the used vocabularies, lineage, and so forth, to be used meaningfully. Similar observations were also made with respect to knowledge extraction [24] and the creation of knowledge graphs and query answering [20] – *raw data* alone is not sufficient. Intuitively, one may assume that this led to a revival of the early work on foundational (top-level) ontologies. By that time, however, the Semantic Web landscape had changed due to the availability of hundreds of highly heterogeneous Linked Data sourced from a multitude of domains [29]. In such a setting, where *synthesis* becomes a crucial task, semantic interoperability has to be fostered without restricting heterogeneity [21]. This, however, is one of the key strengths of ontology design patterns.

By offering common strategies and building blocks, ontology design patterns act as an interoperability fallback level [4] through which local conceptualizations can differ to a degree required to appropriately model a given domain or application while still sharing a common conceptual core. To give a concrete example, imagine three data providers and their respective ontologies. The first provider offers data about (pedestrian) human mobility captured using smartphones, other mobile devices, and social media. The second provider has a broader view on transportation and offers data about cars, buses, taxis, trucks, and so forth. Finally, the third provider stores sparse GPS-based wildlife tracking data from Californian mountain lions. Further assume that the three used ontologies include the semantic trajectory pattern [18] as a common core component. The pattern models trajectories by fixes and segments between them, i.e., as a scalable discretization of the movement of some entity through space and time. Other than that, the ontologies may vary greatly. For example, the wildlife ontology does not model the *mode of transportation* and all location fixes are derived from GPS collars. In contrast, the pedestrian ontology uses so called *geo-social check-ins* to determine a human’s location, i.e., all fixes are related to some meaningful place, e.g., a restaurant. Finally, the transportation ontology may introduce a type hierarchy for the segments by adding classes such as roads and highways. Clearly, there is some data that cannot be retrieved from all providers via a federated query, e.g., only the pedestrian dataset can be queried for place types. Nonetheless, there may be many cases in which federated queries over these data can be

desirable, e.g., to detect spots where wildlife crosses highways or enters human settlements. Such queries remain indeed possible despite major changes in the used ontologies, as they only require the core classes (e.g., fixes, segments, positioning technologies) defined by the trajectory pattern. Summing up, by contributing to semantic interoperability research, knowledge extraction, query answering, and so forth, patterns fill an important gap towards a more sustainable and easier to use Linked Data Web.

So far, we have briefly introduced ontology design patterns, positioned them within the field of Semantic Web research, reviewed potential reasons for their increasing relevance, and discussed a concrete example for their usage in heterogeneous data infrastructures. Next, we will highlight current and future research issues, which will be taken up again in Chapter 9.

Starting with barriers that currently prevent a wider adoption of patterns, one key issue is the lack of an easy to use and up-to-date repository of quality controlled patterns together with a documentation and bundles of patterns that are tested to work seamlessly together [4]. Some of these issues can be addressed by a new version of <http://www.ontologydesignpatterns.org>, while others will require more research such as the combination of patterns into tested bundles/distributions. This also relates to the broader question of how to determine, and ideally measure, the quality of ontologies and ODPs in specific [15, 34]. Finally, many popular ontologies have been developed before patterns played a substantial role, and could be restructured in order to split them into earlier reusable and maintainable patterns. We believe that overcoming these boundaries would lead to a surge in the adoption of patterns for newly published as well as established Linked Datasets.

Looking forward, there is a number of next research steps that deserve further attention. Examples include the development of an ontology design pattern representation language, a set of formal relations between patterns, semantic *shortcuts* and *views* to provide different application-driven perspectives on pattern collections, more robust ways to handle issues of granularity in ODP modeling and usage, large-scale applications and deployments of patterns to modern data infrastructures, further work on pattern-based ontology alignment, knowledge pattern extraction and discovery, and finally the integration of ODPs in major tools such as ontology editors [16, 23].

Bibliography

- [1] C. Alexander. *The Timeless Way of Building*. Oxford Press, 1979.
- [2] S. Bechhofer, I. Buchan, D. De Roure, P. Missier, J. Ainsworth, J. Bhagat, P. Couch, D. Cruickshank, M. Delderfield, I. Dunlop, et al. Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2):599–611, 2013.
- [3] E. Blomqvist, A. Gangemi, and V. Presutti. Experiments on pattern-based ontology design. In *Proceedings of the fifth international conference on Knowledge capture*, pages 41–48. ACM, 2009.
- [4] E. Blomqvist, P. Hitzler, K. Janowicz, A. Krisnadhi, T. Narock, and

- M. Solanki. Considerations regarding ontology design patterns. *Semantic Web*, 7(1):1–7, 2015.
- [5] V. K. Chaudhri, R. Katragadda, J. Shrager, and M. Wessel. Inconsistency monitoring in a large scientific knowledge base. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 66–79. Springer, 2014.
- [6] P. Clark, J. Thompson, and B. Porter. Knowledge Patterns. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 591–600, San Francisco, 2000. Morgan Kaufmann.
- [7] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, V. Huang, K. Janowicz, W. D. Kelsey, D. L. Phuoc, L. Lefort, M. Leggieri, H. Neuhaus, A. Nikolov, K. Page, A. Passant, A. Sheth, and K. Taylor. The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17:25 – 32, 2012.
- [8] M. d’Aquin, A. Schlicht, H. Stuckenschmidt, and M. Sabou. Criteria and evaluation for ontology modularization techniques. In *Modular ontologies*, pages 67–89. Springer, 2009.
- [9] E. Gamma, R. Helm, R. E. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Reading, MA, 1995.
- [10] A. Gangemi. Ontology design patterns for semantic web content. In Y. Gil et al., editors, *Proc. of the 4th Int. Semantic Web Conference, ISWC 2005*, volume 3729 of *LNCS*, pages 262–276. Springer, 2005.
- [11] A. Gangemi. Design patterns for legal ontology constructions. *LOAIT*, 2007:65–85, 2007.
- [12] A. Gangemi, C. Catenacci, and M. Battaglia. Inflammation ontology design pattern: an exercise in building a core biomedical ontology with descriptions and situations. *Studies in health technology and informatics*, pages 64–80, 2004.
- [13] A. Gangemi and V. Presutti. Towards a pattern science for the semantic web. *Semantic Web*, 1(1-2):61–68, 2010.
- [14] U. Grenander. *Elements of Pattern Theory*. Johns Hopkins University Press, Baltimore, 1996.
- [15] K. Hammar. *Towards an Ontology Design Pattern Quality Model*. PhD thesis, Linköping University Electronic Press, 2013.
- [16] K. Hammar. Ontology design patterns in WebProtégé. In *14th International Semantic Web Conference (ISWC-2015)*. CEUR Workshop Proceedings, 2015.
- [17] D. C. Hay. *Data Model Patterns*. Dorset House Publishing, 1996.
- [18] Y. Hu, K. Janowicz, D. Carral, S. Scheider, W. Kuhn, G. Berg-Cross, P. Hitzler, M. Dean, and D. Kolas. A geo-ontology design pattern for semantic trajectories. In *Spatial Information Theory*, pages 438–456. Springer, 2013.
- [19] P. Jain, P. Hitzler, P. Z. Yeh, K. Verma, and A. P. Sheth. Linked data is merely more data. In *AAAI Spring Symposium: linked data meets artificial intelligence*, volume 11, 2010.

- [20] K. Janowicz and P. Hitzler. Thoughts on the complex relation between linked data, semantic annotations, and ontologies. In *Proceedings of the sixth international workshop on Exploiting semantic annotations in information retrieval*, pages 41–44. ACM, 2013.
- [21] K. Janowicz, F. van Harmelen, J. A. Hendler, and P. Hitzler. Why the data train needs semantic rails. *AI Magazine*, 36(1):5–14, 2015.
- [22] A. G. Nuzzolese, A. Gangemi, V. Presutti, and P. Ciancarini. Encyclopedic knowledge patterns from Wikipedia links. In *International Semantic Web Conference*, pages 520–536. Springer, 2011.
- [23] V. Presutti, E. Daga, A. Gangemi, and E. Blomqvist. eXtreme design with content ontology design patterns. In *Proceedings of the 2009 International Conference on Ontology Patterns-Volume 516*, pages 83–97. CEUR-WS. org, 2009.
- [24] V. Presutti, F. Draicchio, and A. Gangemi. Knowledge extraction based on discourse representation theory and linguistic frames. In *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management, EKAW’12*, pages 114–129. Springer-Verlag, 2012.
- [25] V. Presutti and A. Gangemi. Content ontology design patterns as practical building blocks for web ontologies. In *Conceptual Modeling-ER 2008*, pages 128–141. Springer, 2008.
- [26] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, H. Wang, and C. Wroe. OWL pizzas: Practical experience of teaching OWL-DL: Common errors & common patterns. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 63–81. Springer, 2004.
- [27] V. Rodriguez-Doncel, A. A. Krisnadhi, P. Hitzler, M. Cheatham, N. Karima, and R. Amini. Pattern-based linked data publication: The linked chess dataset case. In *Proceedings of the 6th International Workshop on Consuming Linked Data co-located with 14th International Semantic Web Conference (ISWC 2105), Bethlehem, Pennsylvania, US, 2015*.
- [28] F. Scharffe, O. Zamazal, and D. Fensel. Ontology alignment design patterns. *Knowledge and information systems*, 40(1):1–28, 2014.
- [29] M. Schmachtenberg, C. Bizer, and H. Paulheim. Adoption of the linked data best practices in different topical domains. In *The semantic web-ISWC 2014*, pages 245–260. Springer, 2014.
- [30] J. Tidwell. *Designing Interfaces: Patterns for Effective Interaction Design*. O’Reilly, April 2007.
- [31] W. Van Der Aalst, A. Ter Hofstede, B. Kiepuszewski, and A. Barros. Workflow Patterns. *Distributed and Parallel Databases*, 14:5–51, 2003.
- [32] W. R. Van Hage, V. Malaisé, R. Segers, L. Hollink, and G. Schreiber. Design and use of the Simple Event Model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2):128–136, 2011.
- [33] D. Vrandečić and A. Gangemi. Unit tests for ontologies. In *OTM Confederated International Conferences On the Move to Meaningful Internet Systems*, pages 1012–1020. Springer, 2006.
- [34] D. Vrandečić and Y. Sure. How to design better ontology metrics. In *European Semantic Web Conference*, pages 311–325. Springer, 2007.