Providing MOF-based Domain-Specific Languages with UML Notation

Orlando Avila-García, Antonio Estévez García

Open Canarias, S.L.
C/. Elías Ramos González, 4 - Oficina 304
38001 Santa Cruz de Tenerife, Spain
{orlando,aestevez}@opencanarias.com
http://www.opencanarias.com

Resumen Domain-Specific Languages (DSL) are used to describe software and systems in given domains of interest, providing meaningful language constructs that capture the domain's concerns. There are two main approaches to implement a DSL: customizing UML (Unified Modeling Language) with *profiles* or defining it from scratch by using a metamodeling language such as MOF (Meta-Object Facility). The first approach allows reuse of modelling tools and syntax, while the second one provides more semantically meaningful language constructs specially tailored to their domains. Some studies have proposed a synergetic approach based on the use of UML as notation for MOF-based DSLs. Inspired by that idea, we propose a new synergy, which benefits from the advantages of both approaches, eliminates their drawbacks, and tackles the so-called fragmentation problem.

1. Introduction

Model-Driven Engineering (MDE) [5,27] proposes the use of models as first class entities in software and system development, corresponding to artifacts used and operations (tasks or transformations) performed on them during software development. Models are specified in terms of the so-called Domain-Specific Languages (DSL), formal languages specially tailored to capture different concerns of the software process; this enables its automation through a chain of model transformations until we obtain the source code of the software product.

There are two main approaches to implement a DSL. One first approach is to use a general-purpose modeling language such as UML (Unified Modeling Language) and customizing it to model a given concern by means of *profiles*. This was the approach adopted by early MDA (Model Driven Architecture) [24] tools. A second approach is to define a DSL from scratch by using a metamodeling language such as MOF (Meta-Object Facility) [25]. UML Profiles allows reuse of modeling tools and syntax. However, it suffers from lack of flexibility when trying to implement languages whose syntax and semantics deviate a certain degree from those of UML [13,6,30]. MOF provides more semantically meaningful

languages specially tailored to model given domain's concerns. Some of its known disadvantages are the need for specific tooling and extra syntax design effort.

In its initial form, MDA proposed a conceptual framework to separate business specific concerns from platform-independent and platform-specific ones. This implies a chain of vertical exogenous transformations [23] to develop the system. However, recent undertakings in the MDE community recognize the need for more powerful solutions, second-generation MDE platforms devised to facilitate the separation of the software process in many more concerns, different stakeholder views, and different development states of composite complex systems [4]. Each of these concerns must be addressed with its own well-defined and precisely focused Domain Specific Language. Moreover, other kinds of model relationships beyond vertical exogenous transformations are required, e.g., horizontal (exogenous and endogenous [23]) ones.

A typical example is the treatment of *security* separately from the rest of concerns. This is possible by defining a DSL specifically tailored to describe such a concern at the required level of abstraction. Instances of such a DSL specify, at the given abstraction level, the *security* features of the system.

Due to the rising number of DSLs getting involved in a MDE-based software process, the so-called *fragmentation problem* has arisen [4]: coordinating the interplay of the various kinds of models. This model coordination problem may take a number of forms, and it is an important research issue for second-generation MDE platforms [4].

In this paper we argue that the infrastructure behind the implementation of DSLs seems to play a principal role in the *fragmentation problem*. We will show how some proposed synergies between UML and MOF approaches try (at least implicitly) to tackle that problem. Our contribution is a new synergetic approach that offers a steadier solution; it is based on separating the abstract from concrete syntax of a DSL. We present our approach through a case study in a model-driven reverse engineering project, where multiple concerns are managed minimizing the coordination problem between them.

Sect. 2 explains the use of Domain-Specific Languages (DSL) to perform separation of concerns in MDE. Sect. 3 introduces MOF and UML Profiles approaches to define a DSL. Sect. 4 proposes our synergetic approach. Finally, Sect. 5 draws some conclusions.

2. Separation of Concerns in MDE

Recent conceptions of MDA [5] and other approaches to MDE [15,27] advocate a separation of concerns not only in terms of multiple levels of abstraction, but also by modeling multiple concerns at any of those levels. Each of these concerns must be addressed with its own well defined and precisely focused independent Domain Specific Language (DSL). This gives rise to other kinds of model relationships beyond vertical exogenous transformations, e.g., horizontal (exogenous and endogenous [23]) ones. Moreover, transformations with more

than one input and output models are required, e.g., model weavings and model compositions [10,3].

However, the rising number of DSLs getting involved in a MDE-based software process results in the so-called *fragmentation problem*, an important research issue for second-generation MDE platforms [4]: how to coordinate all these various DSLs managing the interplay of the various kinds of models, and to integrate the different stakeholder's concerns by composing and refining them. This model coordination problem may take a number of forms that need to be investigated. In this paper we propose a technique to carry out separation of concerns which do not derive in coordination problems. This approach is based on the utilization of a explicit distinction between the abstract and concrete syntax of a DSL.

2.1. Distinguishing Abstract Syntax from Concrete Syntax

The MDE community has already recognized the advantages of making a distinction between a DSL's concrete syntax and abstract syntax [8,16,21]. This separation offers higher levels of flexibility, since the way a language instance is visualized can be made independent from the way it is serialized (stored in a XML file) or maintained in memory while manipulated by model transformations or program generators. Moreover, it enables the use of other notations apart from the diagrammatical one, e.g., textual notation [22,17].

This distinction just seems to follow the approach taken in compiler theory, where it is common to distinguish between *concrete syntax tree* (also known as *parse tree*) and *abstract syntax tree* when parsing a text-based language [2]. The language's concrete syntax is normally specified through a formal grammar defined in BNF (Bakus-Naur Form). This separation facilitates the development of compilers, as it is easier to translate—to, e.g. machine code—from the abstract structure of the program.

Another example is the separation between content and presentation in XML (eXtensible Markup Language) [18]. Data logical structure is defined through a XML Schema or a DTD (Document Type Definition), while the layout is specified by applying external CSS (Cascading Style Sheets) or XSL (Extensible Stylesheet Language) style sheets to XML documents. This separation helps to maintain XML data "clean" from metadata concerning presentation, which in turn enables the use of more than one viewer to present the same piece of data.

Here we state that by utilizing an explicit distinction between abstract and concrete syntax we can enhance the separation of concerns in a MDE architecture. We can create different perspectives or views of the same model by applying different notations (concrete syntaxes), projecting different aspects or facets of the same model on different presentations. Our solution is based on a synergy between the two main approaches to implement a DSL. But before proceeding to describe it in Sect. 4, in the following section we discuss some of the known advantages and disadvantages of both approaches, and why a synergetic approach is desirable.

3. UML Profiles vs MOF

MDE proposes two main approaches to specify and use Domain-Specific Languages (DSL). In the first place, we can take a general-purpose modeling language like UML (Unified Modeling Language) and customize it to fit a given domain. Alternatively, we might want to define a new language by using a metamodeling language such as MOF (Meta-Object Facility) [25], specially designed to fit the specific characteristics of the domain.

Preliminary studies state that both alternatives have their advantages and disadvantages; depending on the problem at hand, one it is better than the other and vice versa [12,6]. We might say that, unless there is a real need to deviate from the UML metamodel, the benefits of using UML Profiles outweigh their limitations. However, in such situations where the required DSL (its concrete syntax, abstract syntax and semantics) differs notably from UML, there is a great chance that a MOF-based or a synergetic approach is the most cost-efficient; in following sections we show the rationale behind this heuristic. Firstly, we discuss some of the problems arisen when trying to customize UML as well as when building a MOF-based DSL from scratch. Later, we introduce some synergetic approaches proposed in the literature, and how they try to cope with the fragmentation problem. Finally, we propose our new synergetic approach, that eliminates many of the drawbacks from the previous approaches.

3.1. The UML Profiles Approach

UML 2.0 provides two main alternatives to get customized [6]: heavy-weight extensions by package merge and light-weight extensions through profiles. The first alternative is based on UML metamodel extensions through merge operations. Among the disadvantages of this technique, we can consider the complexity involved in getting UML syntax extended through merge operations, and the requirement of modifying existing UML editors to support the merged extensions [6]. Therefore, this technique is seldom used and only recommended in those cases where high level of control on the customization is required.

UML Profiles approach provides a set of mechanisms (*stereotypes*, *tagged values*, and *constraints*) for specializing UML elements, by providing them with additional semantics and attributes as well as imposing new restrictions on them, while preserving the UML metamodel and leaving the original semantics of the rest of UML elements unchanged [12]. This technique is specially advantageous because we can use the same UML editor for multiple DSLs; by loading the corresponding profile, the editor would ideally get customized.

Among the disadvantages of this technique we can consider the obligation to take all UML's large and complex syntax, which may bring problems to users and tool developers [11]. Possibly due to this fact, most commercial UML tools fail, in practice, to give full support to the "strict" application of profiles; that is, although extensions through stereotypes and tagged values are usually supported, validation of the corresponding OCL constrains is seldom satisfactory

[30] and the tools all too often fail to restrict the number of available UML elements. This lack of "strictness" complicates the work of the user when creating an instance of the DSL defined through a profile, because the tool fails to get properly customized (tool bars, contextual menus, etc.).

Another important issue is the questionable universality of the UML [29]. In some circumstances, we might want to use a different visual notation, e.g., in case of domains where experts already share a non-UML notation [15], or architectural views that do not follow the object-oriented paradigm [13]. Moreover, some studies claim that, for some levels of abstraction and concerns, functional, logic and procedural textual languages have proven to be the most expressive solutions [29]. Although UML Profiles proposes a mechanism to customize the diagrammatical notation of UML, this mechanism is quite poor, and suddenly not suitable for defining textual notations.

Finally, a rising experience in MDE practitioners is pointing out difficulties when performing model transformations using DSLs defined through profiles. This is due to the way model elements are typed. Using a UML profile, the semantics of standard UML elements is enriched by labeling them with stereotypes and complementing them with tagged values. Thus, a model element within the UML profile approach will be conforming to a UML metaclass, while its domain-specific meaning will be expressed through a relationship with a decorator containing its stereotypes and tagged values. This mechanism might give rise to semantic checking problems that difficult model-driven tasks [20], such as edition, visualization and transformation.

3.2. The MOF Approach

MOF provides more semantically meaningful languages specially tailored to given domains. In this way, the new language's syntax can be defined to fit the specific characteristics of the domain [30].

This technique offers an implicit separation of concrete and abstract syntax; the metamodel specifies the latter [8,21] while the former must be defined using any other formalism, such as BNF (Backus-Naur form) for textual notations. Usually, we prefer to use a diagrammatical notation to visualize our DSL instances; some studies have proposed the use of UML as concrete syntax for MOF-based DSLs [26,14,7]—this is the case of many of the synergetic approaches explained in the following section.

The difficulties mentioned above, about performing model transformations through UML profiles, are opposed by the simplicity implied in MOF-based DSLs. Notice that a model conforming to a MOF-based DSL is made of elements that conform to a certain metaclass, which possess domain-specific semantics and specific attributes. This typing mechanism is better to carry out model-driven development than typing through other model elements (such as stereotypes) [20]. This has resulted in the preference, from the part of MDE practitioners, to have models conforming to MOF-based DSLs when automated model analysis (e.g., consistency and integrity checking) and code generation is required [14,30].

In a recent comparison between UML Profiles versus MOF, the latter outperformed the former in *Expressive power*, *Flexibility*, *Clarity of semantics*, *Simplicity of constraints*, and *Model Notation*, only being *beaten* regarding the *Tool support* issue [30]. Naturally, the advantages of using a specially tailored DSL are obtained at the cost of developing specific tools and extra syntax design effort.

4. UML Profiles and MOF synergies

There have been studies favoring the synergy between UML Profiles and MOF approaches. [12] proposes the use of MOF to define the DSL's abstract syntax as a previous step to create the corresponding profile. This initial step facilitates the specification of the DSL as a UML profile. The following synergetic approaches make use of UML as visual notation for MOF-based DSLs.

4.1. Unidirectional mappings

[14] proposes the use of UML as notation to visualize models conforming to a given MOF-based DSL. This is carried out through an unidirectional mapping between those models and UML models with profiles. This $\mathtt{MOF} \to \mathtt{UML}$ Profile transformation is done just for documentation purposes, i.e., to provide DSLs with a diagrammatical concrete syntax engineers are accustomed to. This approach enables thus the use of general-purpose UML editors to visualize the models (as documentation).

This technique is interesting because it resembles the use of XSLT (Extensible Stylesheet Language Transformations) to carry out on-the-fly transformation of XML data as to fit a specific presentation layout [28], such as HTML or SVG (Scalable Vector Graphics). This transformation is unidirectional too, as this layout is not meant to edit XML content, but just to present it.

4.2. Bidirectional mappings

A bidirectional mapping between UML and MOF-based DSLs has also been proposed [1]. In that study, a bridge metamodel is proposed to specify the correspondence between every element of a MOF-based metamodel and a UML profile. From that specification, two transformations are automatically derived, making up the bidirectional mapping (MOF \rightarrow UML Profile and UML Profile \rightarrow MOF transformations) that will reconcile both worlds. The specification of the UML profile has to be done manually though.

A further effort to bridge UML Profiles and MOF-based DSLs is proposed in [31], where a dedicated bridge language between MOF-based metamodels and the UML metamodel itself is used. This specification enables the automatic derivation of both the UML profile and the transformations making up the bidirectional mapping. In this way, the development of the profile is automatic. Authors point out some limitations in their approach, the most important one

being a loss of information during round-trip due to the fact that "some transformations can be inverted and some transformations can't" [31](pag.6). That is, in some circumstances, MOF \rightarrow UML Profile and UML Profile \rightarrow MOF transformations are not completely symmetric.

The two previous approaches benefit from the reuse of both UML editors and notation. Besides, they enable the use of both UML Profiles and MOF-based approaches at will in a MDE architecture, having both MOF-based and UML-based models syncronized. They tackle thus the fragmentation problem (see Sect. 2) arisen in those situations, at the cost of having to build a generative infrastructure to bridge both worlds.

4.3. MOF-based DSL editors with UML notation

[26] uses UML notation to visualize and edit instances of OASIS (Open and Active Specification of Information Systems) in an ad-hoc editor. This DSL is used in a MDA-based production environment, although it is not a MOF-based language. The advantage here is clear: not "reinventing the wheel" and using an intuitive diagrammatical notation based on UML, a standard and "unified graphical notation" among software engineers [26](pag. 30).

A related approach has been proposed in [7], where editors store and manipulate models conforming to MOF-based DSLs while presenting them to the user using UML notation. The metamodel is enriched with annotations that tell the editor what UML notational constructs to use to visualize each of its elements. This mechanism enables the use of a unique general-purpose editor, which is able to inspect the metamodel in order to create the proper UML projection to present and edit its instances.

The great advantage of this approach is the reuse of the same editor to edit any DSL. Among the disadvantages we can consider, firstly, the inconvenience of "dirtying" the metamodel with concrete syntax concerns, which violates the principle of separation between concrete and abstract syntax (see Sect. 2.1); and secondly, the inability to use more than one concrete syntax or notation to present the content of a given model.

4.4. Multi-presentation of MOF-based DSLs using UML notation

In this paper we propose to use UML as notation for MOF-based DSLs. Moreover, our technique seeks to provide multiple presentation layouts for the same DSL—i.e., to provide a given DSL with more than one concrete syntax—based on UML notation. In order to do so, we propose to generate lightweight ad-hoc editors for any aspect of the model that needs to be separately presented and edited. This allows different stakeholders to visualize and edit the same model from different points of view, being presented with specific aspects or facets of the model.

The idea is to develop an ad-hoc editor per presentation layout. The main a-priori drawback of this approach is the cost of developing an editor per concrete syntax; however, recent advances in meta-case environments are starting to offer cheap ways to develop them. This is the case of Eclipse GMF (Graphical Modeling Framework) and Visual Studio DSL Tools, which offer a generative approach to produce editors by specifying the abstract and concrete syntaxes of the DSL. The use of those meta-tools reduce time and effort dramatically to obtain ad-hoc editors, which enables the viability of our approach.

This approach is specially useful for carrying out separation of concerns without coordination problems. A facet or aspect of a given concern (modeled using a DSL) is presented to the stakeholder through an ad-hoc editor, which creates for him a specific presentation of the model content by filtering the desired model elements, and formatting them with specific UML notational elements.

5. Conclusions

In this paper we have proposed a new approach to provide MOF-based DSLs (Domain Specific Languages) with UML notation. It follows a synergetic approach between the two main alternatives to implement a DSL: UML (Unified Modeling Framework) Profiles and MOF (Meta-Object Facility). This approach is based on the utilization of an explicit distinction between the abstract and concrete syntax of a DSL. While the former is defined through a MOF-based metamodel, the latter is described in terms of UML notational constructs. This distinction allows us to specify and develop different presentation layouts for a given DSL, which can be used to filter and show different aspects of itL.

Our approach facilitates the separation of concerns in a MDE architecture, tackling the so-called *fragmentation problem*. Multiple stakeholders can share the same model (in its abstract form), inspecting and editing it through customized views, fitted to the task at hand. Coordination problems are thus eliminated, as editors project a subset of a model content, presenting the user with it without any off-line transformation process. Moreover, these tools allow the edition of the content of the repository through the very same layouts or views used to present it. This facilitates the combined work of the different stakeholders.

In our vision, multiple editors would be generated by using a meta-tool; by means of generative techniques, ad-hoc editors would be produced regarding both abstract and concrete syntaxes of the DSL. Each editor would be created as to filter a specific subset of model elements and present them using UML notational elements.

In software architecture, views "emphasize certain aspects of the system while de-emphasizing or ignoring other aspects, all in the interest of making the problem at hand tractable" [9]. It might be interesting to relate our approach with software architecture description techniques. For example, an interesting refinement of the IEEE1471/2000 [19] recommendations for architectural description has been proposed in [14]. In this study, the concept of metamodel is incorporated in the specification, and an interesting distinction between concern, viewpoint, view, model and diagram is made. An unified definition of those terms would be of special interest, which takes into account both software architecture and MDE practitioners.

6. Acknowledgements

Orlando Avila-García is supported by the *Programa Torres Quevedo* of the *Ministerio de Educación y Ciencia, Gobierno de España* (PTQ05-02-03062).

Referencias

- Anas Abouzahra, Jean Bézivin, Marcos Didonet Del Fabro, and Frédéric Jouault.
 A practical approach to bridging domain specific languages with uml profiles. In In proceedings of Best Practices for Model Driven Software Development, at OOP-SLA'05, Oct 2006.
- Alfred V. Aho, Ravi Sethi, and Jeffrey D. Ullman. Compilers: Principles, Techniques, and Tools. Addison Wesley, 1987.
- 3. Jean Bézivin, Salim Bouzitouna, Marcos Didonet Del Fabro, Marie-Pierre Gervais, Fréderic Jouault, Dimitrios Kolovos, Ivan Kurtev, and Richar F. Paige. A canonical scheme for model composition. In *ECMDA-FA 2006: Proceedings of the Second European Conference on Model-Driven Architecture, Foundations and Applications*, volume 4066 of *LNCS*, pages 346–360. Springer-Verlag, Jul 2006.
- Jean Bézivin, Jean-Marie Favre, and Bernhard Rumpe. First international workshop on global integrated model management. In ICSE '06: Proceeding of the 28th international conference on Software engineering, pages 1026–1027, New York, NY, USA, 2006. ACM.
- 5. Alan W. Brown. Model driven architecture: Principles and practice. Software and Systems Modeling (SoSyM), 3(4):314–327, December 2004.
- 6. James Bruck and Kenn Hussey. Customizing UML: Which technique is right for you? White paper, Eclipse UML Project, Jul 2007.
- 7. Achim D. Brucker and Juergen Doser. Metamodel-based uml notations for domain-specific languages. In ATEM 2007: Proceedings of the 4th International Workshop on Software Language Engineering, at MODELS'07, Oct 2007.
- 8. Tony Clark, Andy Evans, Paul Sammut, and James Williams. Language driven development and MDA. *MDA Journal*, Oct 2004. Electronic journal available at http://www.bptrends.com/.
- 9. Paul Clements, Felix Bachmann, Len Bass, David Garlan, James Ivers, Reed Little, Robert Nord, and Judith Stafford. *Documenting Software Architectures: Views and Beyond*. Addison-Wesley Professional, 2002.
- Thomas Cottenier, Aswin Van Den Berg, and Tzilla Elrad. Modeling aspectoriented compositions. In Proceedings of the 7th International Workshop on Aspect-Oriented Modeling, held in conjunction with MoDELS'05., Oct 2005.
- Robert B. France, Sudipto Ghosh, Trung Dinh-Trong, and Arnor Solberg. Modeldriven development using uml 2.0: Promises and pitfalls. Computer, 39(2):59, 2006.
- 12. Lidia Fuentes-Fernández and Antonio Vallecillo-Moreno. An introduction to UML profiles. *The European Journal for the Informatics Professional*, 5(2):6–13, Apr 2004.
- 13. David Garlan, Shang-Wen Cheng, and Andrew J. Kompanek. Reconciling the needs of architectural description with object-modeling notations. *Sci. Comput. Program.*, 44(1):23–49, 2002.

- 14. Bas Graaf and Arie van Deursen. Visualisation of domain-specific modelling languages using uml. In ECBS '07: Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems, pages 586–595, Washington, DC, USA, 2007. IEEE Computer Society.
- 15. Jack Greenfield and Keith Short. Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. John Wiley and Sons, 2004.
- 16. Jack Greenfield and Keith Short. Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools, chapter 8 Language Anatomy. John Wiley and Sons, 2004.
- 17. Hans Grönniger, Holger Krahn, Bernhard Rumpe, Martin Schindler, and Steven Völkel. Text-based modeling. In ATEM 2007: Proceedings of the 4th International Workshop on Software Language Engineering, at MODELS'07, Oct 2007.
- 18. David Hunter, Andrew Watt, Jeff Rafter, Jon Duckett, Danny Ayers, Nicholas Chase, Joe Fawcett, Tom Gaven, and Bill Patterson. *Beginning XML*. Wrox, 3rd edition, 2004.
- IEEE. Recommended practice for architectural description of software intensive systems. Std. 1471-2000, IEEE, 2000.
- 20. Thomas Khune. Making modeling languages fit for model-driven development. In ATEM 2007: Proceedings of the 4th International Workshop on Software Language Engineering, at MODELS'07, Oct 2007.
- 21. Anneke Kleppe. A language description is more than a metamodel. In ATEM 2007: Proceedings of the 4th International Workshop on (Software) Language Engineering, at MODELS'07, Oct 2007.
- 22. Anneke Kleppe. Towards the generation of a text-based IDE from a language metamodel. In *ECMDA-FA 2007: Third European Conference on Model-Driven Architecture, Foundations and Applications*, volume 4530 of *LNCS*, pages 114–129. Berlin Heidelberg, 2007.
- Tom Mens, Krzysztof Czarnecki, and Pieter Van Gorp. A taxonomy of model transformations. Dagstuhl Seminar Proceedings 04101, 2005. Available at http://drops.dagstuhl.de/opus/volltexte/2005/11.
- OMG. MDA guide version 1.0.1. Final Adopted Specification omg/2003-06-01,
 OMG, Jun 2003. Available at http://www.omg.org/docs/omg/03-06-01.pdf.
- 25. OMG. Meta Object Facility (MOF) 2.0 core specification. Final Adopted Specification ptc/2004-10-15, OMG, Apr 2004. Available at http://www.omg.org/docs/ptc/04-10-15.pdf.
- 26. Oscar Pastor and Juan Carlos Molina. *Model-Driven Architecture in Practice: A Software Production Environment Based on Conceptual Modeling.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.
- 27. Douglas C. Schmidt. Model-Driven Engineering. Computer, 39(2):25–31, 2006.
- 28. Jeni Tennison. Beginning XSLT 2.0: From Novice to Professional. Apress, 2nd edition, 2005.
- 29. Dave Thomas. UML unified or universal modeling language? *Journal of Object Technology*, 2(1):7–12, Jan/Feb 2003.
- 30. Ingo Weisemöeller and Andy Schüerr. A comparison of standard compliant ways to define domain specific languages. In ATEM 2007: Proceedings of the 4th International Workshop on Software Language Engineering, at MODELS'07, Oct 2007.
- 31. Manuel Wimmer, Andrea Schauerhuber, Michael Strommer, Wieland Schwinger, and Gerti Kappel. A semi-automatic approach for bridging DSLs with UML. In *Proceedings of the 7th OOPSLA Workshop on Domain-Specific Modeling*, Oct 2007.