Evolution in Model-Driven Engineering

Louis M. Rose

May 21, 2009

Abstract

To be completed.

This report comprises XX words, as counted by detex | wc -w.

Contents

1 Introduction

Software evolution is a development activity in which a software system is updated in response to some external pressure (such as changing requirements). [Brooks 1986] observes that engineering increasingly complicated systems with traditional development approaches presents many challenges including a resistance to evolution. Software evolution is discussed in Section 1.1.

Model-Driven Architecture (MDA) is a software engineering framework defined by the Object Management Group (OMG) [OMG 2008]. MDA provides a set of standards for developing computer systems in a model-centric, or *model-driven*, fashion. In *model-driven engineering* (MDE), models are utilised as the primary software development artefact. Several approaches to MDE are prevalent today, such as [Stahl *et al.* 2006], [Kelly & Tolvanen 2008] and [Greenfield *et al.* 2004]. The approaches vary in the extent to which they follow the standards set out by MDA.

In MDE, the primary software development artefacts are *models*. A model is a description of a phenomenon of interest [Jackson 1996], and may have either a textual or graphical representation. A model provides an abstraction over an object, which enables engineers of differing disciplines to reason about that object [Kolovos *et al.* 2006]. Employing the MDA guidelines has been shown to significantly improve developer productivity and the correctness of software [Watson 2008].

1.1 Software Evolution in Model-Driven Engineering

In the past, studies [Erlikh 2000, Moad 1990] have found that the evolution of software can account for as much as 90% of a development budget; there is no reason to believe that the situation is different now. [Sjøberg 1993] identifies reasons for software evolution, which include addressing changing requirements, adapting to new technologies, and architectural restructuring.

Software development often involves constructing a system by combining numerous types of artefact (such as source and object code, build scripts, documentation and configuration settings). Artefacts depend on each other. Some examples of these dependencies from traditional development include: compiling object code from source code, generating documentation from source code, and deploying object code using a build script. When one artefact is changed, the development team updates the other artefacts accordingly. Here, this activity is termed *migration*.

MDE introduces additional challenges for controlling and managing evolution [Mens & Demeyer 2007]. For example, MDE prescribes automated transformation from models to code. Transformation may be partial or complete; and may take a model direct to code or use intermediate models [Kleppe *et al.* 2003]. Any code or intermediate models generated by these transformations are interdependent with other development artefacts – e.g. a change to a model may have an impact on other models [Deursen *et al.* 2007]. The process of maintaining consistency between interdependent models is termed *model synchronisation*.

To specify a transformation for a model, the model must have a well-defined set of structural elements and rules that it must obey. This information can be specified in a structured artefact termed a *metamodel*. A model is dependent on its metamodel. When a metamodel is changed, its models may require migration. The process of maintaining consistency between a model and its metamodel is termed *(model and metamodel) co-evolution*.

1.2 Research Aim

In traditional software development, some migration activities can be automated (e.g. background incremental compilation of source code to object code), while some must be performed manually (e.g. updating design documents after adding a new feature). MDA seeks to reduce the amount of manual migration required to develop software, but presently no tools for MDE fully support automated evolution; managing evolution in the context of MDE remains an open research problem.

The aim of our research is to develop structures and processes for managing evolutionary changes in the context of Model-Driven Engineering. In our progress report, we presented and discussed data from existing MDE projects. The analysis of this data led us to conclude that little data was available for studying model synchronisation. Consequently, we decided to focus on developing structures and processes for managing (model and metamodel) co-evolutionary changes.

2 Proposed Thesis Structure

We now discuss the structure of the proposed thesis, highlighting completed and remaining work.

2.1 Introduction

The introduction will be based on the introduction and literature review chapters of my qualifying dissertation. However, as Model-Driven Engineering (MDE) is an emerging discipline, any content taken from my dissertation will be updated.

2.1.1 Model-Driven Engineering

The proposed thesis will begin by discussing the challenges that MDE addresses. Subsequently, terminology relevant to MDE will be introduced (including terms such as *model*, *metamodel* and *model transformation*). The benefits of MDE will be discussed, along with the main threats to its adoption. These threats include the additional challenges for controlling and managing software evolution with MDE [Mens & Demeyer 2007].

2.1.2 Software Evolution

The introduction will then discuss software evolution, and its causes. The challenges presented by software evolution will be highlighted, particularly in the context of MDE, and used to motivate the proposed thesis.

2.1.3 Research Aim

The high-level aim of the research will be stated, providing a context for the background and literature review chapters.

2.1.4 Research Method

This section will discuss the way in which the research was conducted, including a discussion of the evaluation strategy.

2.2 Background

The background chapter will serve two purposes: Firstly, to introduce areas of computer science that are related to our research, and secondly to introduce two categories of evolution observed in model-driven engineering. These two categories were described in my progress report.

Again, the background section will be based partly on the literature review section of my qualifying dissertation.

2.2.1 Related Areas

Several subsections will be used, one per related area. Topics are likely to include domain-specific languages and language-oriented programming; refactoring and design patterns; and iterative and incremental approaches to software engineering. We will discuss the applicability and relationship of each area to software evolution in the context of model-driven engineering.

2.2.2 Categories of Evolution in MDE

This section will discuss model and metamodel co-evolution and model synchronisation, two categories of evolution observed in MDE. These categories were introduced in Section 1 of this thesis outline.

2.3 Literature Review

The literature review chapter will provide a thorough review and critical analysis of software evolution research. We will compare and contrast existing techniques for managing and automating activities relating to software evolution. As well as reviewing techniques that apply to the specific challenges caused by software evolution in the context of MDE, we will also critique literature from related areas, such as database and XML schema evolution; and program and modelling language evolution. This chapter will conclude by providing high-level research objectives in the context of the reviewed literature.

2.4 Analysis

The literature review will motivate a deeper analysis of existing techniques for managing evolution in the context of MDE. The benefits and drawbacks of existing techniques will be highlighted by applying them to data from projects using MDE. The analysis will be used to identify requirements for our research.

2.4.1 Locating Data

The first section of the analysis chapter will be based on a section of my progress report, which discusses the data (existing MDE projects) used to analysis existing techniques for managing evolution in the context of MDE. We will introduce and explain the requirements on the data to be used for analysis, identify candidate MDE projects, describe the selection process, and provide reasons for our choices.

2.4.2 Analysing Existing Techniques

Having described the selection of suitable data for the analysis, we will then outline the way in which we have applied existing techniques to the data, and introduce criteria against which the effectiveness of existing techniques will be measured. This work has now been completed, and is discussed in Section **??** of this thesis outline.

2.4.3 Requirements Identification

The analysis of existing techniques will lead to requirements for our research. We will conclude the chapter by enumerating these requirements, refining the high-level research objectives from the literature review chapter into lower-level research objectives.

2.5 Implementation

The implementation chapter will describe the way in which we have approached the requirements presented in the analysis chapter. The requirements will be fulfilled by implementing several related solutions. The solutions will take different forms, including domain-specific languages, automation, and extensions to existing modelling technologies.

2.5.1 Metamodel-Independent Syntax

XMI, an OMG standard for metamodel interchange, and EMF, arguably the most widely used modelling framework, serialise models in a metamodel-specific manner. Consequently, information from the metamodel is required during deserialisation. If the metamodel has evolved since the model was last serialised, deserialisation may fail. This limitation has a major impact on existing techniques for performing co-evolution, forcing them to store original and evolved copies of metamodels.

In a submission to ASE 2009, we have highlighted the problems that a metamodel-specific syntax poses for managing and automating co-evolution, and described our solutions. We prescribe the use of a metamodel-independent syntax for storing models. We also show other ways in which a metamodel-independent representation can be useful for managing and automating co-evolution: Specifically, checking consistency with any metamodel, and performing automatic consistency checking when a new metamodel version is encountered.

2.5.2 Human-Usable Textual Notation

The Human-Usable Textual Notation is an OMG standard textual concrete syntax for the MOF metamodelling architecture. The notation is metamodel-independent – it can be used with any model that conforms to any MOF-based metamodel. HUTN provides a human-usable means for visualising and specifying models, even when those models are inconsistent with their metamodel.

We have developed the only implementation of OMG HUTN, publishing our work at MoDELS 2008 ([Rose *et al.* 2008a]). We have discussed the way in which HUTN may be used during semi-automated migration of inconsistent models in our submission to ASE 2009.

2.5.3 DSL for Migration Strategies

Some of the requirements presented in the analysis chapter can be addressed with a domain-specific language for specifying migration strategies. Migration strategies will be specified as a model transformation on inconsistent models expressed in a metamodel-independent representation. Using a metamodel-independent representation affords us some advantages over existing techniques (such as being able to store partially consistent models). A domain-specific language, rather than an existing model-to-model transformation language, is required to address the specific requirements of model migration, as discussed in my progress report.

2.5.4 Further solutions

Further solutions will be required to meet all of the requirements outlined in the analysis chapter. These solutions are likely to include:

- Extending the DSL for migration strategies: [Herrmannsdoerfer *et al.* 2008] identify the need for re-usable migration strategies, encoded independently of the evolving metamodel. Supporting re-usable migration strategies will likely involve extending the DSL discussed above.

- A metamodel refactoring browser for EMF: inspired by the Smalltalk refactoring browser, this tool will provide common refactorings made to improve the design of existing metamodels. Refactorings will preserve model and metamodel consistency.

- Model-driven migration: As discussed in our ASE 2009 submission, a metamodel-independent syntax enables new approaches to co-evolution. For instance, co-evolution can be driven from models, rather than their metamodel:

  1. Discover a new concept to be modelled.
  2. Represent an existing model that lacks this concept using a metamodel-independent syntax, and generate corresponding HUTN.
  3. Evolve the model in HUTN to express the new concept.
  4. Check consistency with the existing metamodel. Reconcile any inconsistencies by evolving the metamodel.

5. Use the model evolution from step 3 to guide migration of other models.

We will explore and implement new approaches for automated management of co-evolution, such as the one outlined above.

2.6 Evaluation

The evaluation chapter will outline our evaluation method and results, including the impact and limitations our our research; and discuss the extent to which the requirements identified in the analysis chapter have been fulfilled. Evaluation will be conducted in three ways: application of our structures and processes to a sizeable MDE project; publication of our research in academic journals, international conferences and workshops; and assessing the contribution made when delivering our work through an Eclipse research incubation project.

2.6.1 Case Study

We will apply our structures and processes to the Eclipse Generative Modelling Framework (GMF) project [Gronback 2006]. GMF allows the definition of graphical concrete syntax for metamodels. GMF prescribes a model-driven approach: Users of GMF define concrete syntax as a model, which is used to generate a graphical editor. In fact, five models are used together to define a single editor using GMF.

GMF defines the metamodels for graphical, tooling and mapping definition models; and for generator models. The metamodels have changed considerably during the development of GMF. Some changes have caused inconsistency with GMF models. Presently, migration is encoded in Java. Gronback has stated[1] that the migration code is being ported to QVT (a model-to-model transformation language) as the Java code is difficult to maintain.

We identified GMF as the most appropriate candidate for the analysis phase of our research. Consequently, we decided to reserve GMF for the evaluation of our work.

2.6.2 Publications

Publication in academic journals, and at international conferences and workshops ensure that our work is reviewed by experts, and is well-established and communicated in our field of research. So far, I have been the primary author for publications at one international conference ([Rose *et al.* 2008a]), one European conference ([Rose *et al.* 2008b]), and one workshop ([**?**]). The former was published at MoDELS, which has a typical acceptance rate of 20% out of 250-300 papers, and has been nominated by HISE for the departmental best research student paper award.

Where appropriate, we will aim to publish our work at software evolution conferences, as well as at model-driven engineering conferences. Doing so will allow us to assess the impact of our research in a broader sense.

[1] Private communication, 2008.

2.6.3 Delivery through Eclipse

The tools produced as part of our research have been and will continue to be released as part of the Epsilon project, a member of the research incubator for the Eclipse Modeling Project (EMP), arguably the most active MDE community at present. EMP's research incubator hosts a limited number of participants, selected through a rigorous process. Contributions made to the incubator undergo regular technical review.

Contributing to Epsilon allows us to deliver our research to the growing community [?] of Epsilon users.

2.6.4 Limitations

We will discuss the limitations of our work, using where appropriate feedback from users, reviews of publications and scenarios from the case study for context.

2.7 Conclusion

The conclusion will provide a summary of the challenges addressed by and the objectives of our research. We will summarise the way in which we have approached the challenges and met the objectives, concluding with a summary of the evaluation and discussion of future work.

3 Progress

3.1 Achievements

We now reproduce the goals described in my progress report, and discuss the achievements made since submission of that report.

Plan stake-holder survey No existing co-evolution research identifies requirements from developers working on MDE projects. By surveying developers working on existing MDE projects, I planned to ascertain data which would be used to derive requirements for my research. The survey would find answers to the following types of questions: Which tools are developers using for editing and versioning their models and metamodels? Are developers regularly introducing inconsistencies between their models and metamodels? Are developers performing co-evolution manually or using a tool? Which tools are being used for co-evolution?

Goals: Devise and conduct a survey of developers working on existing MDE projects. Identify a process for devising an effective survey. Determine suitable questions, and use the answers to derive requirements for my thesis.

Progress: To discuss conducting a survey of developers, I met with Chris Power and Paul Cairns (members of the York Human Computer Interaction group, who both have experience in developing surveys). Power and Cairns advised me not to conduct a survey, because producing a concise survey containing clear, unambiguous language would be very difficult. Instead, I should interview experts in the field to ascertain the most widely used and unambiguous terms, and then conduct a survey. However, because I only have a limited amount of time to dedicate to requirements analysis, I have decided not to pursue surveying developers. Instead, I will focus on analysing existing techniques and collaborating with colleagues working with evolving metamodels.

Analyse COPE and Cicchetti's work Both [?] and [?] describe tools for performing co-evolution. By analysing both tools with data located from existing MDE projects, I have continued to identify areas in which these tools are effective, and ways in which they may be improved. The analysis has provided requirements for my research.

Goals: Use the example data discussed in my progress report to determine the effectiveness and shortcomings of existing tools for performing automated co-evolution. Use the findings to derive requirements for my research.

Progress: I devised six experiments for the co-evolution tools identified above. Each experiment explored different capabilities of the tools. We have used the findings to identify and motivate requirements for our work. For example, one of the experiments assessed the effectiveness of the tools when managing the migration of a small number of inconsistent models. Considerable effort was required to use either of the tools for any amount of automated migration. Consequently, I felt that the tools provided diminishing returns when used to manage co-evolution in the face of a small number of inconsistent models. In our submission to ASE 2009, we discuss this situation in more detail, describe a semi-automated solution, and provide a concrete example.

Collaborate with Barber and with Sampson I have continued to collaborate with Barber and with Sampson to iteratively and incrementally produce metamodels as discussed in my progress report. Initially, I will collect a record of evolutionary changes made during the development of metamodels. If we encounter any evolutionary changes that inhibit development, I will be able to derive further requirements for my research.

Goals: Determine the extent to which the development of Barber's and Sampson's metamodels will aid my research. Observe and record any evolutionary changes made during the development. Obtain requirements from the data, and from Barber's and Sampson's experiences with MDE.

Progress: Collaboration with Barber has allowed me to observe several evolutionary changes, two of which I had not previously observed for any other metamodel. Work with Sampson is ongoing. One of Sampson's colleagues, Jon Simpson, a research student, will be further developing Sampson's metamodel, and developing model management operations for that metamodel. I am optimistic that Simpson's work will produce data for an initial study of model synchronisation, a category of evolutionary change in MDE for which I presently have no data.

Plan metamodel evolution language Before starting any development, I have consolidated the results of previous activities to produce requirements for a co-evolution language. In addition, I have begun to prototype the language. The primary aim of the prototype will be for me to gain experience with any unfamiliar technologies.

Goals: Produce a list of requirements for a co-evolution language. Investigate any unfamiliar technologies that may aid in the development of the language.

Progress: I have produced the following list of requirements for the DSL for migration strategies:

1. TODO

Presently, I am investigating implementation strategies. I will likely implement the DSL as an internal extension to an existing language, such as Ruby or Scala. Another option is to contribute a new language to Epsilon, a model management framework providing a re-usable architecture for specifying DSLs that manipulate EMF-based models.

Write paper for MoDELS / SLE / MCCM 2009 The research conducted before July 2009 has yielded publishable results. In my progress report, I stated that the collaboration with Barber would be used to generate a report describing our experiences with current MDE tools. We would be able to highlight the need for automated co-evolution tools and discuss why this need is not yet being fulfilled.

Goals: Publish a paper at MoDELS 2009 (or co-located conferences). The paper will provide a basis for a chapter of my thesis.

Progress: The work with Barber (and with Sampson) highlighted deficiencies with existing modelling technologies. In particular, we found several issues that inhibited our ability to develop a metamodel iteratively and incrementally in a collaborative environment. These issues motivated requirements on our research. We have submitted a paper to ASE 2009 that discusses this issues and describes our solutions. We chose ASE rather than MoDELS to assess the contribution in the context of a wider audience (ASE is an software engineering conference, rather than a model-driven engineering conference), and because the deadline for ASE was later than for MoDELS. If the submission is not accepted at ASE, we will revise and send it to another conference.

3.2 Plan

References

[Brooks 1986] Frederick P. Brooks. No silver bullet – essence and accident in software engineering (invited paper). In *Proc. International Federation for Information Processing*, pages 1069–1076, 1986.

[Deursen *et al.* 2007] Arie van Deursen, Eelco Visser, and Jos Warmer. Model-driven software evolution: A research agenda. In *Proc. Workshop on Model-Driven Software Evolution*, pages 41–49, 2007.

[Erlikh 2000] Len Erlikh. Leveraging legacy system dollars for e-business. *IT Professional*, 2(3):17–23, 2000.

[Greenfield *et al.* 2004] Jack Greenfield, Keith Short, Steve Cook, and Stuart Kent. *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools.* Wiley, 2004.

[Gronback 2006] Richard Gronback. Introduction to the Eclipse Graphical Modeling Framework. In *Proc. EclipseCon*, Santa Clara, California, 2006.

[Herrmannsdoerfer *et al.* 2008] Markus Herrmannsdoerfer, Sebastian Benz, and Elmar Juergens. Automatability of coupled evolution of metamodels and models in practice. In *Proc. International Conference on Model Driven Engineering Languages and Systems*, volume 5301 of *LNCS*, pages 645–659. Springer, 2008.

[Jackson 1996] Michael Jackson. *Software Requirements and Specifications.* Addison-Wesley and ACM Press, 1996.

[Kelly & Tolvanen 2008] Steven Kelly and Juha-Pekka Tolvanen. *Domain-Specific Modelling.* Wiley, 2008.

[Kleppe *et al.* 2003] Anneke G. Kleppe, Jos Warmer, and Wim Bast. *MDA Explained: The Model Driven Architecture: Practice and Promise.* Addison-Wesley, 2003.

[Kolovos *et al.* 2006] Dimitrios S. Kolovos, Richard F. Paige, and Fiona A.C. Polack. The Epsilon Object Language (EOL). In *Proc. European Conference on Model Driven Architecture – Foundations and Application*, volume 4066 of *LNCS*, pages 128–142. Springer, 2006.

[Mens & Demeyer 2007] Tom Mens and Serge Demeyer. *Software Evolution.* Springer-Verlag, 2007.

[Moad 1990] J Moad. Maintaining the competitive edge. *Datamation*, 36(4):61–66, 1990.

[OMG 2008] OMG. Object Management Group home page [online]. [Accessed 30 June 2008] Available at: http://www.omg.org, 2008.

[Rose *et al.* 2008a] Louis M. Rose, Richard F. Paige, Dimitrios S. Kolovos, and Fiona A.C. Polack. Constructing models with the Human-Usable Textual Notation [to appear]. In *Proc. International Conference on Model Driven Engineering Languages and Systems*, volume 5301 of *LNCS*, pages 249–263. Springer, 2008.

[Rose *et al.* 2008b] Louis M. Rose, Richard F. Paige, Dimitrios S. Kolovos, and Fiona A.C. Polack. The Epsilon Generation Language. In *Proc. European Conference on Model Driven Architecture – Foundations and Applications*, volume 5095 of *LNCS*, pages 1–16. Springer, 2008.

[Sjøberg 1993] Dag I.K. Sjøberg. Quantifying schema evolution. *Information & Software Technology*, 35(1):35–44, 1993.

[Stahl *et al.* 2006] Thomas Stahl, Markus Voelter, and Krzysztof Czarnecki. *Model-Driven Software Development: Technology, Engineering, Management.* Wiley, 2006.

[Watson 2008] Andrew Watson. A brief history of MDA. *Upgrade*, 9(2), 2008.