Integration of Handwritten and Generated Object-Oriented Code

Timo Greifenberg[1], Katrin Hölldobler[1], Carsten Kolassa[1], Markus Look[1],
Pedram Mir Seyed Nazari[1], Klaus Müller[1(✉)], Antonio Navarro Perez[1],
Dimitri Plotnikov[1], Dirk Reiss[2], Alexander Roth[1], Bernhard Rumpe[1],
Martin Schindler[1], and Andreas Wortmann[1]

[1] Software Engineering, RWTH Aachen University, Aachen, Germany
`{greifenberg,hoelldobler,kolassa,look,nazari,mueller,`
`perez,plotnikov,roth,rumpe,schindler,wortmann}@se-rwth.de`
[2] Institute for Building Services and Energy Design, TU Braunschweig,
Braunschweig, Germany
`reiss@sse-tubs.de`

Abstract. In many development projects models are core artifacts used to generate concrete implementations from them. However, for many systems it is impossible or not useful to generate the complete software system from models alone. Hence, developers need mechanisms for integrating generated and handwritten code. Applying such mechanisms without considering their effects can cause issues in projects, where model and code artifacts are essential. Thus, a sound approach for the integration of both forms of code is needed.

In this paper, we provide an overview of mechanisms for integrating handwritten and generated object-oriented code. To compare these mechanisms, we define and apply a set of criteria. The results are intended to help model-driven development (MDD) tool developers in choosing an appropriate integration mechanism. In this extended version, we additionally discuss essential integration aspects including the protection of generated code and elaborate on how to use action languages to extend generated code.

1 Introduction

The vision to create complex software systems from abstract models by systematically transforming these into concrete implementations [1] is pursued by model-driven development (MDD) [2]. The prevailing conjecture, however, is that deriving a non-trivial, complete implementation from models alone is not feasible [3]. Thus, current MDD techniques require tool developers to integrate generated and handwritten code. Various mechanisms can be used to perform this code integration. However, there is no ultimate integration mechanism that should always be used. Instead, it depends on the context in which this code integration has to be carried out and on the concrete requirements which integration mechanisms are best suited to be applied.

We examined existing mechanisms to integrate handwritten and generated object-oriented code to support MDD tool developers in selecting integration mechanisms. Additionally, we created a set of criteria focusing on different properties of code integration mechanisms to assess and compare these mechanisms. The presented criteria are based on a decade of experiences in object-oriented software engineering and MDD research [4,5], code generator development and code integration research [6,7], and experiences with MDD processes within various domains including automotive [8], cloud computing [9], robotics [10], and smart buildings [11].

We introduce eight handwritten code integration mechanisms and evaluate each with respect to our criteria. Their strengths and weaknesses are shown in the evaluation results. By means of this, we seek to increase the comparability between the integration mechanisms. In particular, this overview is intended to be used by MDD tool developers to find a proper integration mechanism on a case-by-case basis.

The list of integration mechanisms and evaluation criteria presented in this paper does not claim to be complete. However, if further integration mechanisms need to be compared or the mechanisms need to be evaluated with respect to additional evaluation criteria, this paper can be used as a basis easily.

In summary, the contributions of this paper are:

– A list of evaluation criteria for code integration mechanisms.
– A collection of mechanisms to integrate generated and handwritten code.
– An evaluation of the integration mechanisms based on the list of evaluation criteria.

In previous work, we have already presented different mechanisms to integrate handwritten and generated code for object-oriented programming languages [12]. This article is an extension of our previous work. It presents action languages as a further approach to customize generated code in addition to integrating handwritten code. Additionally, we provide an extended discussion on protecting generated code from being overridden.

The remainder of this article is structured as follows. We begin by introducing criteria to assess the code integration mechanisms (Sect. 2). The mechanisms are separated into mechanisms based on specific concepts in programming languages (Sect. 3) and mechanisms free of such requirements (Sect. 4). Subsequently, we summarize and discuss the evaluation results (Sect. 5). After that we elaborate on related work (Sect. 6) before we briefly present action languages as an alternative to extend generated code. Finally, we conclude this contribution (Sect. 8).

2   Evaluation Criteria

The following criteria concern different properties of integration mechanisms and aim at helping developers in selecting a suitable mechanism. These criteria are based on existing literature [13,14], best practices in software development [15,16], and our experience [4–11]. This list does not claim to cover all

aspects of handwritten code integration. Nonetheless, it can be used as an initial list of criteria which can be adapted to fit personal needs. The presented criteria are not weighted on purpose since a weighting is highly subjective and also tailored to an application scenario, which is not intended.

C1: Separation of Generated and Handwritten Code. - Can generated and handwritten code be separated into different files?

Separation of concerns is an essential design practice in software development [15] and has been proposed as a criterion to evaluate integration mechanisms for handwritten code by [13,14]. One crucial benefit of separating generated and handwritten code into different files is that it can be ensured that the generator does not overwrite handwritten code. In case of mixing generated and handwritten code in one file, the handwritten code might not always be preserved.

C2: Support for Overriding Generated Parts. - Can developers add handwritten parts that are used instead of particular generated parts?

Depending on the developer's requirements it can be necessary to adapt particular parts of the generated functionality. This can be done by integrating handwritten code that refines these parts. A benefit of such handwritten code refinements is that the code generator does not have to be changed to fit different requirements.

C3: Extendability of the Generated Interfaces. - Can the generated interfaces be extended with handwritten methods?

Hiding implementation details is accepted as common practice in software development [15,16]. Accordingly, we assume that functionality of the generated system is provided to developers through dedicated generated interfaces and that the system's functionality is only accessed by using these. Obviously, these generated interfaces are oblivious to handwritten code. Consequently, the generated interfaces need to be extended to allow access to handwritten functionality.

C4: Independence of Handwritten Code at Generation-time. - Is the generator independent of the existence of handwritten code at generation-time?

In some handwritten code integration mechanisms the generated code needs to be adapted if handwritten code is present. In this case the handwritten code is processed by the generator and the generated code is adapted accordingly. For instance, the handwritten code is merged into the generated code and one artifact is produced. If this functionality is not provided by the generator framework, a generator developer has to extend the generator with such functionality. This additional effort might not be desired and can be avoided by choosing a generator framework with support for handling handwritten code. However, then the choice of an integration mechanism influences the choice of the generator framework. This is not always feasible.

C5: Independence of Additional OOP Language Constructs. - Can the integration mechanism be realized using only default OOP language constructs?

Some of the existing handwritten code integration mechanisms are tailored to a particular type of OOP language that provides specific language constructs.

Consequently, such integration mechanisms are restricted to generators that generate code in one of these languages. The benefit of handwritten integration mechanisms using default OOP language constructs is that no additional tooling is required and the generator is not tailored to a specific type of language. The following language constructs are regarded as the default OOP language constructs in this work: (abstract) classes, inheritance, interfaces, object creation facility and message-passing capability. Except for interfaces this understanding complies with [17]. We are aware that not all OOP languages provide the concept of interfaces but interfaces can be realized using classes with empty method bodies and inheritance.

3   Integration Mechanisms Based on Language Concepts

In this section, we present a catalog of integration mechanisms that presuppose certain concepts in the target language, for instance, inheritance known from object-oriented programming. Each presented mechanism is described and evaluated with the following scenario:

Assume the input model for the code generator is an UML class diagram (CD) containing the class Editor. As CDs do not model class behavior, implementations of Editor methods can be developed manually and the resulting handwritten code needs to be integrated with the code generated for Editor.

The same problems arise whenever modeling languages do not support modeling of all aspects of a system and these parts have to be developed manually. The mechanisms presented in the remainder of this publication are not limited to CDs and a particular type of code generation. We only use CDs to give an illustrative example.

3.1   Generation Gap

The generation gap mechanism [13,18,19] assumes that an interface and a default implementation are generated for each class in the input model. For instance, the interface Editor and the default implementation EditorBaseImpl are generated for the class Editor. Manual extensions of specific methods or behavior different from the default implementation are defined in the handwritten class EditorImpl. Figure 1 depicts this pattern for the class EditorImpl. Please note, that here and in the following «gc» denotes generated code and «hc» denotes handwritten code.
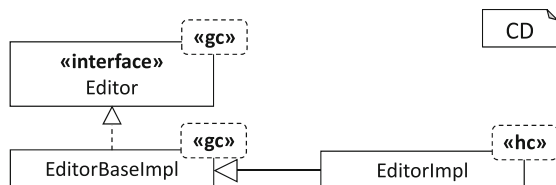

Fig. 1. Generation gap pattern for the Editor example.

In this case, `EditorImpl` is the implementation that will be used by both the generated code as well as manually written code that uses the interface `Editor`.

Please note, that the generation gap mechanism requires developers to create the handwritten class, no matter whether handwritten code is inserted into that class or not. In projects in which handwritten extensions are rarely needed, this leads to bloated projects with an unnecessary high number of artifacts.

Evaluation

C1: Separation of Generated and Handwritten Code. Fulfilled. This criterion holds by definition of the pattern, as the handwritten code has to be stored in separate classes.

C2: Support for Overriding Generated Parts. Fulfilled. The possibility to override generated methods is a crucial feature of the pattern.
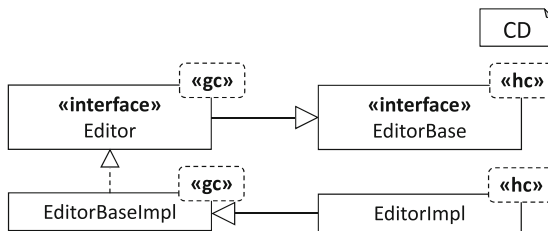
C3: Extendability of the Generated Interfaces. Unfulfilled. This approach does not provide means to reflect added methods in the generated interface. The extended generation gap mechanism (see Sect. 3.2) addresses this issue.

C4: Independence of Handwritten Code at Generation-time. Fulfilled. Whether or what handwritten code exists does not influence the code generation in any way.

C5: Independence of Additional OOP Language Constructs. Fulfilled. This mechanism does not require additional OOP language constructs.

3.2   Extended Generation Gap

A mechanism that addresses two disadvantages of the basic generation gap mechanism (see Sect. 3.1) - the inability to extend the generated interface and the necessity to create an implementation class - is the extended generation gap mechanism, as shown in Fig. 2. Since this mechanism has been developed for our particular needs, its name is not a well-known term in MDD.

Fig. 2. Extended generation gap pattern with an additional handwritten Interface.

The first disadvantage is addressed by allowing to add a handwritten interface on top of the generated interface. As the generated interface `Editor` extends the handwritten interface `EditorBase`, all methods added to `EditorBase` are also available when accessing `Editor`. However, developers do not have to add this

handwritten interface. Instead, the generator checks at generation-time whether it exists. If it does exist, the generated interface will extend the handwritten interface. Consequently, the generator needs to be executed again after adding a handwritten interface to reflect this change in the generated code.

When a developer adds a handwritten interface, the handwritten implementation class (`EditorImpl` in Fig. 2) has to be provided as well. If no handwritten interface is present, the generator generates a concrete class `EditorBaseImpl` by default and an additional implementation class does not have to be added by developers. In this way, developers are not forced to integrate their own implementation class. However, if a developer adds the handwritten class `EditorImpl`, which has to extend the generated base class `EditorBaseImpl`, this class is used in the generated code and `EditorBaseImpl` becomes abstract. This integration of handwritten code requires the generator to be executed again, as it checks at generation-time whether developers added their own implementation classes.

Other variations of the generation gap mechanism are possible. For instance, assuming that a handwritten interface always exists. A detailed discussion is neglected because the variations are very similar and, as shown in the example, differ in technical details.

Evaluation

C1: Separation of Generated and Handwritten Code. Fulfilled. See Sect. 3.1.

**C2: Support for Overriding Generated Parts.** Fulfilled. See Sect. 3.1.

C3: Extendability of the Generated Interfaces. Fulfilled. The API of the generated class can be extended easily by adding a handwritten interface which is extended by the generated interface. Thus, method signatures which are added to the handwritten interface are also available in the generated one. The actual implementations of these methods have to be added to the handwritten implementation class.

C4: Independence of Handwritten Code at Generation-time. Unfulfilled. The generator has to check whether a handwritten interface or implementation class was introduced, as this influences the structure of the generated code.

C5: Independence of Additional OOP Language Constructs. Fulfilled. See Sect. 3.1.

3.3   Delegation

Delegation is a pattern of object composition in object-oriented programming. In essence, the pattern consists of two objects taking the roles of one delegator and one delegate, respectively. The delegator delegates parts of its functionality to the delegate by invoking methods of the delegate. To this end, the delegate provides an interface declaring the method signatures that can be invoked. Figure 3 gives an overview of the objects and relationships involved. Here, `Editor` is the delegator and `EditorDelegateImpl` is the delegate implementing the methods defined in the `EditorDelegate` interface.
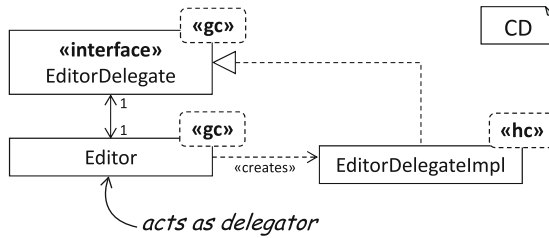
Fig. 3. Delegation pattern requires a delegator for regarding handwritten implementations.

The delegator is responsible for instantiating the delegate. `EditorDelegate-Impl` is the implementation that will be used by both the generated code as well as manually written code that uses the interface `EditorDelegate` in this case.

In essence, the purpose of delegation is to outsource functionality to a distinct object with an explicit interface specific to this functionality. This purpose makes delegation naturally applicable to the integration of handwritten and generated code. The roles of the delegators are taken by generated classes while the roles of the delegates are taken by handwritten classes. All functionality that cannot be generated is delegated to the handwritten delegates. The delegate interfaces, thus, are well-defined and distinct interfaces between generated and handwritten code. It can be generic and handwritten or specific and generated. The choice depends on whether the delegated functionality depends on the model or not. For instance, to delegate the implementation of method signatures in class diagrams to a delegate, it is appropriate to generate the delegate interface based on the method signatures defined by the CD.

In general, delegation provides a higher degree of encapsulation and cohesion compared to alternative patterns. Moreover, it avoids inheritance in handwritten classes since delegates only have to implement an interface. In programming languages without support for multiple inheritance delegation allows developers to use inheritance with handwritten classes.

Evaluation

C1: Separation of Generated and Handwritten Code. Fulfilled. The pattern separates generated and handwritten code by putting them into different classes and interfaces.

C2: Support for Overriding Generated Parts. Unfulfilled. In this mechanism, only designated delegators can be implemented to provide handwritten code. It is not possible to override other generated parts.

C3: Extendability of the Generated Interfaces. Unfulfilled. The generated interface can be extended by subinterfaces and concrete delegators according for the extended subinterface. However, the generated delegator is not aware of these extensions.

C4: Independence of Handwritten Code at Generation-time. Fulfilled. The existence of handwritten delegate classes does not influence the code generation.

C5: Independence of Additional OOP Language Constructs. Fulfilled. The default OOP language constructs suffice to implement this mechanism.

Alternatives. The cardinalities of the delegation relationship are not necessarily restricted. Thus, an implementation of the pattern may associate one delegator with exactly one delegate, or one delegator with many delegates, or many delegators with one delegate. The choice between these variants depends heavily on whether the delegate is stateful or stateless. Stateless delegates can generally be shared by many delegators and do not need to be instantiated repeatedly.

3.4 Include Mechanism

Include mechanisms are based on dedicated language constructs which allow to define that a certain file should be included into another file at a specific point. This idea can be easily used to integrate generated and handwritten files as either a generated file includes handwritten files (see Fig. 4) at designated places or vice versa. In general, the effect of using an include statement is equivalent to injecting the content of the included file to the corresponding location in the including file. Specific languages may offer include mechanisms with different meanings, but this will not be discussed in the following as the focus is on the general idea of include mechanisms.

By including handwritten files in generated files, the generator can define the required structure of the files and developers merely need to introduce selected handwritten files, which are included properly without the developer having to worry about it. This is advantageous if developers should not be able to deviate from this generated structure, as they can only provide the handwritten files which are included. Thus, developers are guided in which files to provide. On the other hand, if developers need more flexibility and should be able to deviate from such a generated structure, including generated files in handwritten files is more appropriate. This variant is, of course, accompanied by the risk that developers forget to include the proper generated files at the proper places.
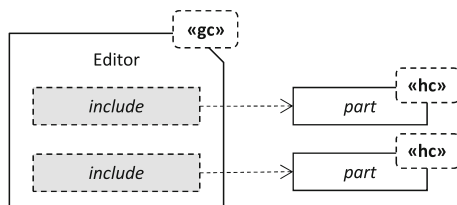
Fig. 4. The include mechanism adds include statements to the generated file to consider handwritten artifacts.

Evaluation

C1: Separation of Generated and Handwritten Code. Fulfilled. No matter whether generated files include handwritten files or vice versa, generated and handwritten parts are separated into different files.

C2: Support for Overriding Generated Parts. Unfulfilled. It is not possible to integrate handwritten code which is used instead of generated code as handwritten code can only be included and, therefore, added.

C3: Extendability of the Generated Interfaces. Conditionally fulfilled. To fulfill this criterion, a programming language has to allow include statements inside of interfaces to extend the signature. Even though we are not aware of a language that supports this, the concept itself does not forbid it. Therefore this criterion might be fulfilled, depending on the specific target language.

C4: Independence of Handwritten Code at Generation-time. Fulfilled. The generation of the include functionality does not depend on the existence of handwritten code.

C5: Independence of Additional OOP Language Constructs. Unfulfilled. The mechanism requires include constructs which do not belong to the default OOP language constructs.

3.5   Partial Classes

Partial classes facilitate splitting class implementations into several source code files. These parts are merged in a pre-compilation phase. The result contains the union of all methods, fields and super types of all its partial definitions. In contrast to aspect-oriented programming (see Sect. 3.6), partial classes are concerned with only one class rather than multiple. In Fig. 5 this approach is illustrated.

The partial classes mechanism suits well for integrating handwritten and generated code. Each generated partial class can be extended by adding handwritten code in its own partial class in a separate source file. The resulting generated and handwritten code is integrated automatically by merging them. This merging can either be done by applying naming conventions, i.e., partial classes with the same name are merged, or by explicit notations. How and which partial classes are merged is defined by the used language.

The CD in Fig. 5 illustrates the partial class mechanism. In this case, the generated code is stored in the partial class `EditorBaseImpl` and the handwritten code is stored in a separate partial class `EditorBaseImpl`.

Evaluation

C1: Separation of Generated and Handwritten Code. Fulfilled. The handwritten and generated partial classes can be located in different source code files.

C2: Support for Overriding Generated Parts. Conditionally fulfilled. In general, the partial classes mechanism does not forbid to override methods' implementations. However, depending on the used programming language that supports partial classes, this criterion may not be fulfilled.

C3: Extendability of the Generated Interfaces. Conditionally fulfilled. The concept of the partial classes mechanism can be applied to interfaces, too.
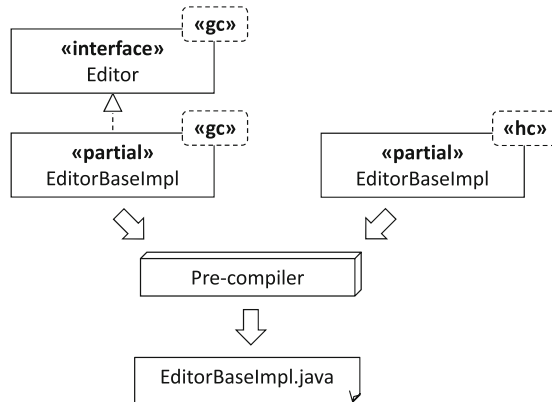
Fig. 5. Partial classes mechanism merges the handwritten and the generated implementation to one single artifact.

Thus, the additional method signatures can be added to the handwritten partial interface which is merged with the generated partial interface. However, if a realization of the partial class mechanism supports partial classes but not partial interfaces, this criterion is not fulfilled.

C4: Independence of Handwritten Code at Generation-time. Fulfilled. Handwritten code does not have to exist at generation-time, because its existence does not influence the code generation process. Handwritten code only has to be available when the pre-compiler merges the generated and handwritten code.

C5: Independence of Additional OOP Language Constructs. Unfulfilled. This mechanism requires support for partial classes which is not regarded as a default OOP language construct in this work.

3.6 Aspect-Oriented Programming

Aspect-oriented programming (AOP) [20] addresses crosscutting concerns - functionality or features scattered across several classes causing duplication - by encapsulating duplicated code in one place. Although integrating handwritten code with generated code does not necessarily deal with crosscutting concerns, AOP can be used for this integration [7]. One advantage in this context is that the generated code does not need to offer a specific architecture to be extendable by handwritten code. Instead, the handwritten code is added by so called *aspects*, as shown in Fig. 6.

An aspect reacts to a predefined event (*pointcut*) during the program execution. Such a predefined event can be, for instance, a method call of a specific method in a specific class. The action that is executed when a pointcut is reached is implemented in an *advice*. Such an advice can be executed before, after or instead of the according event.
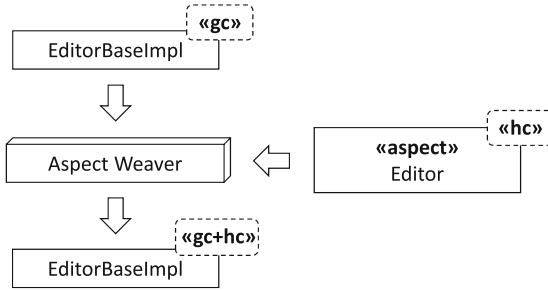
Fig. 6. Overview of an AOP integration mechanism for a part of a generated software system.

The integration of handwritten code can be performed by implementing an advice that is executed instead of, before or after a specific generated method. All these cases, of course, require the generator to create at least a dummy implementation of the corresponding method so that the handwritten advice can be executed instead of that generated method.

Figure 6 illustrates the idea underlying the integration of handwritten code using AOP. In this case, handwritten code is added to an advice in an aspect. An aspect weaver then takes the instructions given in the aspect and produces a combined artifact (e.g. a source code file) in which the advice implementation is woven into the code of the generated classes. This means that the advice instructions are introduced into the proper locations in the generated classes. In the example given in Fig. 6, the implementation of a generated method in `EditorBaseImpl` would be replaced by the advice implementation, if the aspect contains one advice for one method.

Besides the additional overhead of weaving the aspects into the source code, a major drawback of AOP is that it is more difficult to understand the program flow as it is influenced by aspects. Moreover, refactorings in the source code may lead to invalid aspects, known as the fragile pointcut problem [21].

Evaluation

C1: Separation of Generated and Handwritten Code. Fulfilled. AOP offers a clear separation of the handwritten and the generated code. The generated code is not aware of the aspects, which contain the handwritten code and which are stored in separate files.

C2: Support for Overriding Generated Parts. Fulfilled. As described above, an advice can be implemented such that it is called instead of a particular event in the generated class. By means of this, the execution of a generated method can be prevented and instead the handwritten implementation is executed.

C3: Extendability of the Generated Interfaces. Fulfilled. Concepts in AOP allow to extend interfaces and classes. Consequently, the API can be extended.

C4: Independence of Handwritten Code at Generation-time. Fulfilled. The generator and the generated code is not aware of handwritten code at all.

C5: Independence of Additional OOP Language Constructs. Unfulfilled. In order to be able to use this mechanism, the generated code needs to conform to a programming language that supports AOP or contains aspect-oriented extensions. This is not provided by OOP languages by default.

Alternatives. If the target language does not support aspect-orientation, hook points can be created in the generated code. Every hook point is called at the beginning and end of a method execution. By using inheritance, these hook points can be used to add behavior before or after the actual method execution. The behavior can be changed completely by overriding the method representing the hook point (see generation gap in Sect. 3.1). To some extent, this mechanism simulates aspects in AOP but it is not able to extend the API.

4   General Integration Mechanisms

Besides integration mechanisms that rely on language concepts, integration approaches free of this restriction are presented in this section. In compliance with Sect. 3, these approaches are evaluated with respect to the criteria described in Sect. 2.

4.1   PartMerger Mechanism

A PartMerger is a component that is capable of merging multiple files of a specific type, e.g., Java files, into one file. Obviously, this idea fits well to integrate handwritten and generated parts, as these parts can be separated into different files and later be merged by the PartMerger as shown in Fig. 7.
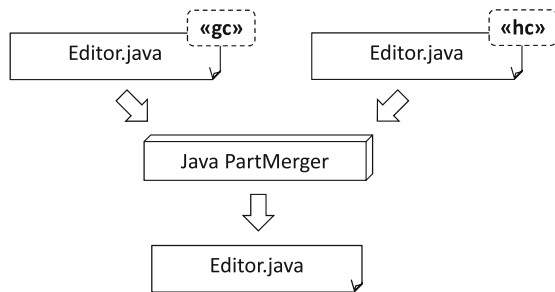
Fig. 7. The PartMerger mechanism merges source code artifacts (e.g. Java source code) to one artifact.

The PartMerger mechanism is a generalization of the partial classes mechanism (see Sect. 3.5). In contrast to partial classes, the PartMerger can also

deal with non-source code artifacts. For instance, the DSL tool bench Monti-Core [22,23] uses this mechanism to merge generated and handwritten Eclipse plugin.xml files and Eclipse manifest files.

Without any restriction on how to merge files, the PartMerger mechanism is very flexible. To consider handwritten code, a PartMerger can give higher priorities to handwritten extensions when merging two files. Furthermore, there are different strategies for invoking the PartMerger and for defining the files to be merged. A simple strategy is to invoke the PartMerger automatically for files conforming to a specific naming convention on the artifact level, e.g., files with the same file name in specific folders or files with a common pre- or postfix. Another strategy is to let the developers configure which files should be merged.

A drawback of the PartMerger mechanism is the lack of tool support when integrating handwritten code. Common functionalities such as code completion are not directly available to access parts of the generated code due to the strict separation of the generated and the handwritten source code files. Instead, developers need to implement such tooling on their own. This is an advantage of applying partial classes (see Sect. 3.5). The according tooling does not have to be implemented by the developer but it is already provided.

The PartMerger mechanism is very similar to the partial classes mechanism. The main difference is rooted in the language support for partial classes. In other words, a language that supports partial classes provides concepts to define what language parts are merged. The compiler takes care of the merging. In contrast, the PartMerger approach is based on a dedicated configurable tool that merges different artifacts, e.g. Java source code artifacts. Consequently, the PartMerger mechanism is not tailored to a particular language. However, the realization of the PartMerger might be tailored to particular languages.

Evaluation

C1: Separation of Generated and Handwritten Code. Fulfilled. The separation of artifacts is a precondition for this approach.

C2: Support for Overriding Generated Parts. Fulfilled. A PartMerger component can be implemented in such a way that it assigns a higher priority to handwritten parts so that certain generated parts are substituted by handwritten parts in the merged artifact. In this way, the handwritten code will be executed instead of the generated code.

C3: Extendability of the Generated Interfaces. Fulfilled. Extending the API of a generated interface is easily possible. To accomplish this, a handwritten interface needs to be merged with the generated interface.

C4: Independence of Handwritten Code at Generation-time. Fulfilled. Handwritten artifacts do not have to exist at generation-time, because they do not influence the generation-process. Instead, handwritten artifacts only have to be available when the PartMerger merges the generated and handwritten artifacts. This takes place after the code generator has finished.

C5: Independence of Additional OOP Language Constructs. Fulfilled. This mechanism does not require any kind of OOP language construct at all.

4.2   Protected Regions

Protected regions are designated regions located in generated code that allow to add handwritten code (see Fig. 8). A common use case for applying protected regions is to generate method signatures from input models and to insert protected regions into the corresponding method bodies.

Each protected region is typically surrounded by comments comprising a unique identification string. In this way, it can be differentiated between different protected regions. Before (re)generating code, the generator identifies protected regions in the generated code and manages the code contained in these regions based on the identification strings. While generating code, it reinserts the code previously contained in a particular protected region. As a consequence, the identification string associated with a protected region is crucial to be able to preserve the handwritten code in subsequent generator executions.
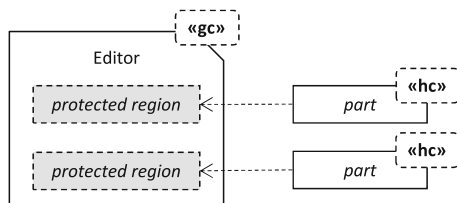
Fig. 8. The protected regions mechanism requires predefined regions that contain handwritten code.

Different model-to-text transformation languages provide built-in support for declaring protected regions, including XPand [24], Acceleo [25], Epsilon Generation Language [26], JET [27] and MOFScript [28]. Some of these languages have different names for the protected region mechanism, e.g., protected area [13], user code block in Acceleo, user region in JET, and unprotected block in MOFScript.

A major drawback of protected regions is that it cannot be guaranteed that the generator preserves handwritten implementations. The reason for this is that handwritten code is mixed with generated code. In addition, to support this mechanism, a guarantee has to be given that the identification string is unique and stable. Otherwise, handwritten code may get lost in some situations.

Evaluation

C1: Separation of Generated and Handwritten Code. Unfulfilled. The handwritten and generated parts are mixed within the same files, therefore there is no separation according to our criterion.

C2: Support for Overriding Generated Parts. Unfulfilled. It is not possible to override generated code. Only explicitly designated parts can be extended.

C3: Extendability of the Generated Interfaces. Fulfilled. An extension of the API can be achieved by generating in such a way that protected regions are introduced into the generated interfaces. Then, methods can be added to that protected region.

C4: Independence of Handwritten Code at Generation-time. Unfulfilled. The generator has to analyze the previously generated code to extract handwritten code from protected regions. Otherwise the generator would not be able to inject that code from the protected regions back into the generated code.

C5: Independence of Additional OOP Language Constructs. Fulfilled. This mechanism does not require any kind of OOP language construct at all.

Alternatives. The Eclipse Modeling Framework (EMF) [29] applies a mechanism to integrate handwritten code which is different to protected regions described so far, but conceptually comparable. In EMF, every class, method etc. that is generated, includes a Javadoc comment that contains a generated tag [30]. By removing the generated tag the generated implementation can be changed. Hence, removing the generated tag corresponds to introducing a protected region.

5   Discussion

In this section, we summarize and discuss the evaluation results shown in Table 1. A plus sign in a table cell indicates that the approach fulfills the corresponding criterion, whereas a minus sign expresses that the criterion was not satisfied. Parentheses denote that the criterion is fulfilled under certain conditions.

All approaches, except for protected regions, separate generated and handwritten code on the basis of files. Thus, it is ensured that the handwritten code is not overwritten because only the generated files are overwritten. However, the protected regions approach combines generated and handwritten code. Consequently, this approach does not protect handwritten code from being overwritten.

Table 1 also shows that the extended generation gap approach, the AOP approach and the PartMerger approach provide the most flexibility when overriding generated parts and extending the generated interfaces. For partial classes, it depends on the actual programming language being used whether it is possible to override generated parts or not. Moreover, it also depends on the programming language, if besides partial classes also partial interfaces are supported. The situation is different for include mechanisms. Here, we are not aware of any language allowing for includes in interfaces to extend the generated interfaces. However, the concept itself does not forbid such behavior. Therefore, its applicability also depends on the used language.

All approaches but the extended generation gap mechanism and protected regions do not need to check for the existence of handwritten code at generation-time. The extended generation gap approach demands the generator to check for the existence of a handwritten interface or implementation class. In case of protected regions, the generator has to extract the handwritten code from the generated code before (re)generating to be able to reinsert it into the generated code.

Table 1. Overview of integration mechanisms and results of analysis with respect to the criteria.

	Generation Gap	Extended Gen. Gap	Delegation	Include Mechanism	Partial Classes	AOP	PartMerger	Protected Regions
C1: Separation of Generated and Handwritten Code	+	+	+	+	+	+	+	-
C2: Support for Overriding Generated Parts	+	+	-	-	(+)	+	+	-
C3: Extendability of the Generated Interfaces	-	+	-	(+)	(+)	+	+	+
C4: Independence of Handwritten Code at Generation-time	+	-	+	+	+	+	+	-
C5: Independence of Additional OOP Language Constructs	+	+	+	-	-	-	+	+

Moreover, Table 1 illustrates that only the following approaches can be used without requiring other language constructs than the default OOP language constructs: generation gap, extended generation gap, delegation, PartMerger and protected regions. All other mechanisms depend on additional language constructs which are not provided by default by OOP languages, e.g., include functionality or partial classes.

A topic that should be discussed when allowing to add handwritten code is the support for restricting what parts of the generated code can be overridden. If it needs to be ensured that specific generated code is executed in the software system, it should not be possible to override such parts. This is mostly true for parts of the software system that define a static and fix behavior without any chance of adaptation. Approaches that do not support overriding generated parts such as delegation, include mechanisms and protected regions ensure this property by construction. This means that handwritten code can only be introduced in designated regions. For generation gap and the extended generation gap, generated methods have to be generated using specific modifiers (e.g. private modifier) of the underlying OOP language to prevent that these methods can be overridden. For the PartMerger approach, it has to be ensured that the PartMerger can be configured in a way to not override generated code. This protected code may be marked in a specific way. When applying AOP, it depends on the programming language used whether it can be restricted that handwritten code is executed instead of specific other generated code. Finally, the partial classes mechanism is generally not designed in a way that parts of the generated code are protected, because the different parts are merged. However, it depends on the compiler and programming language, which may support such a mechanism.

6   Related Work

To the best of our knowledge, no other publication exists which gives a comparable overview and evaluation of different integration mechanisms. However, as most of the presented mechanisms have been described by other authors, we give an overview of existing work in this section.

Pietrek et al. [14] list guidelines on how to integrate generated and handwritten artifacts including generation gap, protected regions, as well as include mechanisms. These mechanisms are also covered in [31].

Similarly, Stahl et al. [13] describe the adaptation of different design patterns - in particular delegation - to integrate handwritten and generated code. These mechanisms are also discussed in [32,33]. The latter covers aspect-oriented methods as well. This concept is also employed in [34–36]. Additionally, a brief overview of integration mechanisms is given in [19]. It includes different variations of the generation gap, partial classes and protected regions mechanisms.

Approaches that specifically target .NET as the target platform and thus allow special language features including partial classes are covered in [37,38]. Both also cover protected regions, as well as generation gap and delegation.

The mechanisms supported in MetaEdit+ [39] are protected regions, functionality externalized in files (similar to the include approach in our paper) as well as the direct inclusion of handwritten code in model files and are described in [40]. Direct inclusion of handwritten code in model files has been left out, because we put the focus on integrating handwritten code in generated code.

The different approaches that affect the implementation of model-driven architecture are presented in [41]. They range from the unidirectional code inclusion to complete round-trip engineering, where code portions of handwritten code are traced and reflected back into the model.

The code generator LLBLGen Pro [42] supports different target infrastructures and allows the integration of handwritten code through protected regions, user-specific templates that are included during the generation process, as well as language-specific features such as partial classes.

The approach described in [43] supports the concept of partial classes for generated object-oriented code and protected regions for code that does not support this mechanism. Additionally, Brückmann et al. [44] advocate patterns such as delegation to incorporate manually written code in generated parts.

## 7   Extend Generated Code with Action Languages

The integration of handwritten code into generated code is mainly done by using mechanisms provided by the target language. A different approach to extend generated code is to embed an action language into the used modeling language.

An action language can be a (restricted) expression language (e.g., a boolean expression language), a DSL for behavior description (e.g., statecharts), or even a general-purpose language such as Java. For instance, UML provides a designated action language [45]. The main idea is to describe the implementation of specific

model parts within the model by using the action language [40]. A code generator is finally responsible for merging this embedded code into the generated code. For instance, the body of a method in a UML class diagram could be described in Java and the generator can merge this Java code into the method body part of the generated code.

The major drawback of this approach is that the models are polluted with technical implementation details. In addition to that, the tool support to implement code within models is typically less sophisticated than the tool support for implementing code in a typical IDE. Mainly, the difference to the integration mechanisms described so far is that the code to be integrated is embedded in the model and not implemented separately in an external source code file. Moreover, the action language may require a code generator to map it to a concrete implementation. Nevertheless, additional code can be added to the generated code quite easily.

There are three variants of applying action languages that are not based on a direct embedding in a modeling language. Figure 9 shows all three variants. The first approach is to embed references to the action language code into the model instead of embedding the complete action language code directly into the model [40]. Even though this decreases the level of pollution of the models, the models are still polluted with technical details. Another variant is that the action language code references the modeling language element where it is to be used. Finally, a configuration can be used to link action language code and the modeling language element. The commonality of all variants is that an infrastructure is required to merge the action language code and the generated code.
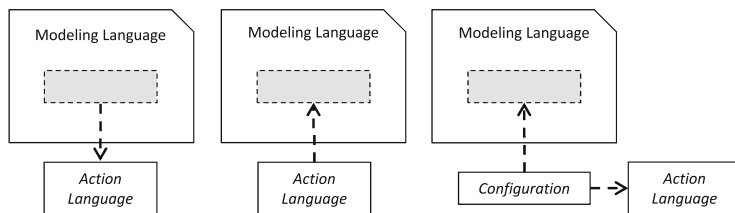
Fig. 9. Three variants of applying action languages.

8   Conclusion

Current model-driven development (MDD) techniques require tool developers to integrate generated and handwritten code to create a complete software system. In this journal paper, we presented eight integration mechanisms useful for OOP languages: generation gap, extended generation gap, delegation, include mechanisms, partial classes, AOP, PartMerger, and protected regions. These mechanisms are compared with respect to a set of five evaluation criteria that aim to address essential aspects of handwritten and generated code integration. For

instance, if generated and handwritten code can be separated or if it is possible to override generated parts. Our evaluation results are summarized in Table 1 and show that no mechanism is the best solution.

Essentially, choosing a suitable mechanism for integrating handwritten and generated code depends on the concrete use case and the associated requirements. The catalog of integration mechanisms and evaluation criteria presented in this paper provides an overview for MDD tool developers that can be used to find an appropriate integration approach on a case-by-case basis. In this extension of our previous work, we present action languages as another approach to integrate handwritten and generated code and discussed issues related to restricting parts that can be overridden to emphasize concerns that have to be regarded when choosing the right mechanism.

The proposed evaluation criteria are not complete and may not be sufficient to choose an appropriate mechanism in special cases. Nevertheless, the approaches and criteria shown in this paper can be used as a foundation that can be adapted and extended to fit specific requirements.

References

1. France, R., Rumpe, B.: Model-driven development of complex software: a research roadmap. In: Future of Software Engineering, ICSE 2007, pp. 37–54. IEEE Computer Society (2007)
2. Kleppe, A.G., Warmer, J., Bast, W.: MDA Explained: The Model Driven Architecture: Practice and Promise. Addison-Wesley Longman Publishing Co., Inc., Boston (2003)
3. Wile, D.S.: Lessons learned from real DSL experiments. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003, pp. 265–290. IEEE Computer Society (2003)
4. Rumpe, B.: Modellierung mit UML, 2nd edn. Springer, Heidelberg (2011)
5. Rumpe, B.: Agile Modellierung mit UML: Codegenerierung, Testfälle, Refactoring. Springer, Heidelberg (2012)
6. Rumpe, B., Schindler, M., Völkel, S., Weisemöller, I.: Generative software development. In: Proceedings of the 32nd International Conference on Software Engineering, ICSE 2010, pp. 473–474. ACM (2010)
7. Schindler, M.: Eine Werkzeuginfrastruktur zur agilen Entwicklung mit der UML/P. Ph.D. thesis, RWTH Aachen University (2012)
8. Grönniger, H., Hartmann, J., Krahn, H., Kriebel, S., Rothhardt, L., Rumpe, B.: Modelling automotive function nets with views for features, variants, and modes. In: Proceedings of Embedded Real Time Software and Systems, ERTS 2008 (2008)
9. Navarro Pérez, A., Rumpe, B.: Modeling cloud architectures as interactive systems. In: 2nd International Workshop on Model-Driven Engineering for High Performance and CLoud computing, MDHPCL 2013, Miami, Florida, CEUR Workshop Proceedings, pp. 15–24 (2013)
10. Ringert, J.O., Rumpe, B., Wortmann, A.: From software architecture structure and behavior modeling to implementations of cyber-physical systems. In: Software Engineering 2013 Workshopband, GI, pp. 155–170. Köllen Druck+Verlag GmbH, Bonn (2013)

11. Kurpick, T., Pinkernell, C., Look, M., Rumpe, B.: Modeling cyber-physical systems: model-driven specification of energy efficient buildings. In: Proceedings of the Modelling of the Physical World Workshop, MOTPW 2012, pp. 2:1–2:6. ACM (2012)

12. Greifenberg, T., Hölldobler, K., Kolassa, C., Look, M., Mir Seyed Nazari, P., Müller, K., Navarro Perez, A., Plotnikov, D., Reiss, D., Roth, A., Rumpe, B., Schindler, M., Wortmann, A.: A comparison of mechanisms for integrating handwritten and generated code for object-oriented programming languages. In: Proceedings of the 3rd International Conference on Model-Driven Engineering and Software Development, Angers, France, pp. 74–85. Scitepress (2015)

13. Stahl, T., Völter, M.: Model-Driven Software Development: Technology, Engineering, Management. Wiley, UK (2006)

14. Pietrek, G., Trompeter, J., Niehues, B., Kamann, T., Holzer, B., Kloss, M., Thoms, K., Beltran, J.C.F., Mork, S.: Modellgetriebene Softwareentwicklung. MDA und MDSD in der Praxis. Entwickler.Press (2007)

15. Parnas, D.L.: On the criteria to be used in decomposing systems into modules. Commun. ACM **15**(12), 1053–1058 (1972)

16. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional, Boston (1995)

17. Eliens, A.: Principles of Object-Oriented Software Development. Addison-Wesley Longman Publishing Co., Inc., Boston (1994)

18. Vlissides, J.: Pattern Hatching: Design Patterns Applied. Addison-Wesley, UK (1998)

19. Fowler, M.: Domain Specific Languages. Addison-Wesley, Boston (2010)

20. Kiczales, G., Lamping, J., Mendhekar, A., Maeda, C., Lopes, C., Loingtier, J.-M., Irwin, J.: Aspect-oriented programming. In: Akşit, M., Matsuoka, S. (eds.) ECOOP 1997. LNCS, vol. 1241, pp. 220–242. Springer, Heidelberg (1997)

21. Kellens, A., Mens, K., Brichau, J., Gybels, K.: Managing the evolution of aspect-oriented software with model-based pointcuts. In: Thomas, D. (ed.) ECOOP 2006. LNCS, vol. 4067, pp. 501–525. Springer, Heidelberg (2006)

22. Grönniger, H., Krahn, H., Rumpe, B., Schindler, M., Völkel, S.: MontiCore: a framework for the development of textual domain specific languages. In: 30th International Conference on Software Engineering, ICSE 2008, pp. 925–926. ACM (2008)

23. Krahn, H., Rumpe, B., Völkel, S.: MontiCore: a framework for compositional development of domain specific languages. Int. J. Softw. Tools Technol. Transf. **12**, 353–372 (2010)

24. XPand website, May 2014. http://www.eclipse.org/modeling/m2t/?project=xpand#xpand. Accessed on 13 May 2015

25. Acceleo website, May 2014. http://www.eclipse.org/acceleo/. Accessed on 13 May 2015

26. Rose, L.M., Paige, R.F., Kolovos, D.S., Polack, F.A.C.: The epsilon generation language. In: Schieferdecker, I., Hartman, A. (eds.) ECMDA-FA 2008. LNCS, vol. 5095, pp. 1–16. Springer, Heidelberg (2008)

27. JET website, May 2014. http://www.eclipse.org/modeling/m2t/?project=jet#jet. Accessed on 13 May 2015

28. Oldevik, J., Neple, T., Grønmo, R., Aagedal, J.Ø., Berre, A.-J.: Toward standardised model to text transformations. In: Hartman, A., Kreische, D. (eds.) ECMDA-FA 2005. LNCS, vol. 3748, pp. 239–253. Springer, Heidelberg (2005)

29. Budinsky, F., Steinberg, D., Merks, E., Ellersick, R., Grose, T.J.: Eclipse Modeling Framework, 2nd edn. Addison-Wesley, Boston (2008)

30. Gronback, R.C.: Eclipse Modeling Project: A Domain-Specific Language (DSL) Toolkit. Addison-Wesley, Boston (2009)
31. Petrasch, R., Meimberg, O.: Model-Driven Architecture: Eine praxisorientierte Einführung in die MDA. Dpunkt Verlag, Heidelberg (2006)
32. Völter, M.: A Catalog of Patterns for Program Generation, Version 1.6, April 2003. http://www.voelter.de/data/pub/ProgramGeneration.pdf. Accessed on 13 May 2015
33. Völter, M., Bettin, J.: Patterns for Model-Driven Software-Development, Version 1.4, May 2004. http://www.voelter.de/data/pub/MDDPatterns.pdf. Accessed on 13 May 2015
34. Groher, I., Voelter, M.: Aspect-oriented model-driven software product line engineering. In: Katz, S., Ossher, H., France, R., Jézéquel, J.-M. (eds.) Transactions on Aspect-Oriented Software Development VI. LNCS, vol. 5560, pp. 111–152. Springer, Heidelberg (2009)
35. Völter, M., Groher, I.: Handling variability in model transformations and generators. In: Proceedings of the 7th OOPSLA Workshop on Domain-Specific Modeling, DSM 2007. ACM (2007)
36. Kang, K.C., Sugumaran, V., Park, S.: Applied Software Product Line Engineering. Auerbach Publications, Boston (2009)
37. Dollard, K.: Code Generation in Microsoft .NET. Apress, Berkley (2004)
38. Greenfield, J., Short, K., Cook, S., Kent, S.: Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. Wiley, New York (2004)
39. Tolvanen, J.P., Kelly, S.: MetaEdit+: defining and using integrated domain-specific modeling languages. In: Proceeding of the 24th ACM SIGPLAN Conference Companion on Object Oriented Programming Systems Languages and Applications, OOPSLA 2009, pp. 819–820. ACM (2009)
40. Kelly, S., Tolvanen, J.P.: Domain-Specific Modeling: Enabling Full Code Generation. Wiley, New York (2008)
41. Frankel, D.S.: Model Driven Architecture: Applying MDA to Enterprise Computing. Wiley, New York (2003)
42. LLBLGen Pro website, May 2014. http://www.llblgen.com/. Accessed on 13 May 2015
43. Warmer, J., Kleppe, A.: Building a flexible software factory using partial domain specific models. In: Proceedings of the 6th OOPSLA Workshop on Domain-Specific Modeling, DSM 2006, pp. 15–22. ACM (2006)
44. Brückmann, T., Gruhn, V.: An architectural blueprint for model driven development and maintenance of business logic for information systems. In: Babar, M.A., Gorton, I. (eds.) ECSA 2010. LNCS, vol. 6285, pp. 53–69. Springer, Heidelberg (2010)
45. Object Management Group website: Concrete Syntax for a UML Action Language: Action Language for Foundational UML (ALF) Version 1.0.1 (2013–09-01), May 2014. http://www.omg.org/spec/ALF/1.0.1/PDF/