Generating Formal Specifications from CASE Repositories

B. Ryan
Generating Formal Specifications from CASE Repositories

Benjamin Ryan
School of Computing, Leeds Metropolitan University
Beckett Park Campus, Leeds LS6 3QS
e-mail: b.ryan@lmu.ac.uk
August 25, 1997

Abstract

A method for formally specifying the generation of formal specifications from a CASE tools' repository is described. The limitations of the approach are identified and summarised. An alternative approach is outlined that addresses an important limitation, that of the original approach being specific to one software development method. The use of a neutral representation to generalise the approach to other software development methods is outlined. A mechanism is outlined that uses the original approach and the neutral representation to enable an incremental approach to the adoption of formal specification techniques. The incremental adoption is supported by the default generation of a formal specification that allows the inclusion of hand written formal specification fragments in a controlled way. The work builds on the established area of method integration by thoroughly investigating one approach to integration and suggesting ways that this could be improved. The techniques presented go some way to providing a framework in which the use of formal specifications techniques can be adopted in an incremental manner.

1 Introduction

Formal specification techniques have developed considerably over the past two decades from being the products of pure research to practical techniques that are used by the software development community to specify and design dependable and correct software systems. However, their penetration has been most successful in the area of safety critical and mission critical systems where software failure is unacceptable. Many of these uses have been forced upon developers by standardisation authorities who mandate the use of formally based techniques [3].

The use of formally based techniques has had less of an impact outside this area though this situation is changing slowly [4]. There have been many reasons proposed (and refuted) [12] as to why formally based techniques have enjoyed less success outside these limited areas.

An approach taken by the research community to address this problem is to integrate formally based techniques with other non-formal 1 methods. These approaches, called Method Integration, have concentrated on integrating the complementary aspects of the formal and non-formal to produce an improved method.

The benefits of combining non-formal methods and formal based techniques have been widely reported [12, 30, 23]. Within the literature there are numerous references to the integration disparate software development techniques to provide an integrated method. Semmens et al. give an overview in [30]. Since this overview was written other work on integrated methods has been published detailing attempts to integrate different types of development techniques and different ways of integrating them [18, 4]. Several authors have proposed different classifications for the approaches taken to Method Integration [9, 28, 19, 10, 15].

One approach to method integration, a generative approach, has been proposed by several authors [29, 8, 7, 14, 15]. This paper reports on an implementation of the generative approach. The implementation has been undertaken to determine the resources required to implement such an approach and to analyse the approach for weaknesses. The

---

1 the term non-formal is used to describe both structured and object oriented development methods which do not use a mathematical notation as the primary description mechanism
An improved generative approach is proposed that address some of the negative aspects that have been identified. This improved approach generalises the generative approach by removing the bias towards one particular CASE implementation. The generalisation is gained by using a neutral representation for meta-model descriptions used by the generative approach. Using facilities of the neutral representation the improved approach also addresses a criticism of generative approaches that use of the formal notation directly by the practitioner is hindered.

The rest of this paper is organised as follows. Section 2 describes a generative approach to method integration detailing the four stages. Section 3 presents an analysis of the generative approach and details the resources required, the quality of the formal specification generated and a number of negative aspects of the approach. Section 4 outlines an improved generative approach that addresses certain criticisms of the original approach. Sections 5 and 6 summarise the conclusions and suggest areas for future work.

1.1 The Automated Teller Machine (ATM): An Example

In this paper an example will be used to give a context to the definitions given. This example is a simple Automated Teller Machine (ATM). The ATM system comprises of a number of accounts that are held by customers. There may be more than one customer per account and a customer may hold many accounts. The accounts are divided into two types; Current and Deposit. Current accounts have an overdraft limit and a weekly withdrawal limit, Deposit accounts have an interest rate. The interest is paid on the balance of the account at the end of each month. There are a number of ATMs in the system that may be used by customers who have been issued with cards and pin numbers for these cards. Transactions that customers may make at an ATM are withdraw cash, pay in cash, query a balance, change the PIN of a card and request a statement of previous transactions. The full example is given in [25]. The Entity Relationship Diagram for the ATM example is given in Figure 1.

2 A Generative Approach

The generative approach consists of four stages.


2. A definition of the rules for a well formed meta-model instance that ensures that the generation produces a valid formal specification.

3. A description of the abstract syntax of the target formal language that will be the receiving structures used by the mapping functions.

4. A definition of the functions which form the mapping between meta-model components and receiving structures.

A full treatment of these stages is given in [26].

2.1 Meta-model Descriptions

The initial stage of the formalism is to specify the components of the IE meta-model that are used in the mapping from the non-formal notations to the formal notations. This gives a description of the inputs to the mapping functions that will define the formal semantics of the data, process and behaviour models used in the IE method.

The specification is split into the static and dynamic components of the meta-model. The static components are the artefacts produced from the High Level Data Modelling and the Data Analysis phases of the IE method. The dynamic aspects are the artefacts produced by the High Level Functional Modelling and the Process Analysis phases.

For reasons of space only the meta-model descriptions for the static components Data Types, Attributes and Entities are given. Full descriptions of the other meta-model components are given in [26].
2.2 Static Property Models

Within the IE meta-model there are components used to specify the data structures of interest in the system. These components specify the static properties of the system and serve to classify and name objects of interest. They use standard techniques such Entity Relationship Modelling, Data Flows and, Data Stores.

2.2.1 Data Types

The type of an attribute is its domain and is recorded as a name. A domain may be a PLAIN domain where its name represents a type or an ENUMERATED domain where its name represents a type with a fixed set of values specified by the literals.

\[ \text{DOMAINTYPE} ::= \text{PLAIN} \mid \text{ENUMERATED} \]
2.2.2 Attributes

Attributes are used to describe entities. An attribute definition consists of a unique name and a domain type that has been previously specified.

Instances of attributes which occur within entities have additional properties to plain attributes. To model these additional properties it is necessary to introduce definitions of the properties and their allowable values. These definitions will be reused in the definition of relationships.

An attribute may have a cardinality of one or many when occurring within an entity, this is to allow for Repeating Groups which will be removed in the Normalisation process.

\[\text{CARDINALITY} ::= \text{ONE} \mid \text{MANY}\]

An attribute may be optional or mandatory. If it is mandatory it must have a defined value in all instances of the entity type. Where an attribute is optional it may not be used as an identifying attribute.

\[\text{OPTIONALITY} ::= \text{OPTIONAL} \mid \text{MANDATORY}\]

An attribute can be an identifying attribute and it serves, possibly with other attributes to uniquely identify an entity instance.

\[\text{IDENTIFYING} ::= \text{YES} \mid \text{NO}\]

2.2.3 Entities

Entities are any object that the business wishes to record information about. They have a unique name and may have attributes. The attributes are defined as a finite set of attribute instances and this set is partitioned into two subsets. There are those attributes which identify an entity and those that do not. For an attribute to be an identifying attribute it must not be optional, it must have cardinality one and have been defined as an identifying attribute.
Where an entity owns a relationship that is defined as identifying the identifying attributes of the entity that is the destination of the relationship will be used, in addition to the source entity’s own identifying attributes, to fully identify the source entity. This may result in a progression through a hierarchy of entities and identifying relationships to fully define the set of identifying attributes for a particular entity.

An entity may also be classed as associative or as a subtype of an existing entity. Where it is an associative entity it is used to resolve many-to-many relations and does not require any attributes in addition to the identifying attributes of those entities that it associates.

\[
\text{ENTITYCLASS} ::= \text{PLAIN} \mid \text{ASSOCIATIVE} \mid \text{SUBTYPE}
\]

\[
\text{Entity}
\begin{align*}
\text{name} & : \text{NAME} \\
\text{type} & : \text{ENTITYCLASS} \\
\text{supertype} & : \text{NAME} \\
\text{attributes} & : F \text{AttrInst} \\
\text{identifiers} & : F \text{AttrInst} \\
\text{nonIdentifiers} & : F \text{AttrInst}
\end{align*}
\]

\[
\forall a_1, a_2 : \text{attributes} \mid a_1 \neq a_2 \Rightarrow a_1.\text{name} \neq a_2.\text{name} \\
\langle\text{identifiers, nonIdentifiers}\rangle \text{partition attributes} \\
\text{name} \neq \text{supertype} \\
\text{type} \neq \text{ASSOCIATIVE} \Rightarrow \text{attributes} \neq \emptyset
\]

### 2.2.4 Repository Definition

Assuming similar definitions for the other meta-model components a Repository can now be defined. An Entity Model View (EMV) is the necessary information that an activity has available to it to perform its function. The EMV is made up of three components:

1. an Import View, Import, detailing inputs to an activity,
2. an Export View, Export, detailing the outputs produced by an activity and
3. an Access View, Access, detailing the components of the Data Dictionary affected by the activity.

Within each set of meta-model components there is a restriction that they be uniquely named.
Generating Formal Specifications from CASE Repositories

2.3 Target Language Specification

In a generative approach it is necessary to specify what elements of the specification language, in this case Z, are to be used. These definitions are based upon those used by Semmens and Allen in their work in formalising Yourdon [30] with extensions to include the differences in IE.

This description is necessary to allow the definition of mathematical functions which map meta-model components onto an abstract syntax of a formal language. When this definition is given in a formal manner the syntax and type checking facilities provided by the Z language can be used to ensure the function definition is syntactically correct and that each use of the functions conforms to its typing constraints. This rigorous specification and checking provides a sound basis for the later implementation of the mapping functions in an executable language.

The target language specification appears in the Appendix.

2.4 Mapping Functions

These functions map components from the formal description of the meta-model to the receiving structures specified for the target language. They are specified as partial functions which have their domain specified by set comprehension to ensure that they are well-formed. This is necessary as the functions use meta-model components in their definitions to which the function may not be applied. As an example the function GPlainEntSch below restricts the domain to the entities found in the data dictionary which have class PLAIN and have at least one attribute.

```plaintext
| rep : Repository

| domains : F Domain
| attributes : F Attribute
| entities : F Entity
| relationships : F Relationship
| subjectAreas : F SubjectArea
| entityRelationshipDiagrams : F Erd
| dataFlows : F DataFlow
| dataStores : F DataStore
| externalAgents : F ExternalAgent
| entityModelViews : F EntityModelView
| events : F Event

∀ d1, d2 : domains | d1 ≠ d2 • d1.name ≠ d2.name
∀ a : attributes • a.domain ∈ domains
∀ e1, e2 : entities | e1 ≠ e2 • e1.name ≠ e2.name
∀ rel1, rel2 : relationships | rel1 ≠ rel2 • rel1.name ≠ rel2.name
∀ s1, s2 : subjectAreas | s1 ≠ s2 • s1.name ≠ s2.name
∀ erd1, erd2 : entityRelationshipDiagrams | erd1 ≠ erd2 • erd1.name ≠ erd2.name
∀ df1, df2 : dataFlows | df1 ≠ df2 • df1.name ≠ df2.name
∀ ds1, ds2 : dataStores | ds1 ≠ ds2 • ds1.name ≠ ds2.name
∀ ex1, ex2 : externalAgents | ex1 ≠ ex2 • ex1.name ≠ ex2.name
∀ env1, env2 : EntityModelView | env1 ≠ env2 • env1.name ≠ env2.name
∀ ev1, ev2 : events | ev1 ≠ ev2 • ev1.name ≠ ev2.name
∀ ent : entities • sup : Entity | ent.supertype = sup.name •
  #ent.attributes > #sup.attributest
```

Second BCS-FACS Northern Formal Methods Workshop  6
All functions have the basic form of $funcName : \text{metaModelComponent} \leftrightarrow \text{targetLanguageComponent}$. All functions operate on components of a repository.

It is not possible to detail all the functions for the IE meta-model for reasons of space. The full definitions are given in [26]. For the purposes of explanation the functions relating to Domains, Attributes and Entities are given.

It should be noted that functions used within the definitions are of two types; those ending with the suffix \text{ASF} (Abstract Syntax Function) are the injective functions from the abstract syntax definition of the target formal language. Those beginning with a capital \text{G} (Generation) are mapping functions for meta-model components.

As an aid to comprehension the result of each function is shown below the function. These results are taken from case study produced to evaluate the process of generating a formal specification.

### 2.4.1 Domains

For each domain within the dictionary a \text{Given Set} is declared if the domain is not specified as an enumerated type. If it is enumerated a \text{Free Type} definition will be used.

\begin{align*}
G\text{PlainDomain} : \text{Domain} & \leftrightarrow \text{TYPE} \\
\text{dom} & \in G\text{PlainDomain} = \{\text{dom} : \text{rep.domains} \mid \text{dom.literals} = \emptyset\} \\
G\text{PlainDomain} & = (\lambda d : \text{Domain} \bullet \text{GivenASF}(d.\text{name})) \\
\end{align*}

Accounts in the ATM system have a unique identifying number. The structure of the number is not important at this point.

\begin{align*}
[\text{ACCOUNTNO}] \\
G\text{EnumDomain} : \text{Domain} & \leftrightarrow \text{TYPE} \\
\text{dom} & \in G\text{EnumDomain} = \{\text{dom} : \text{rep.domains} \mid \text{dom.literals} \neq \emptyset\} \\
G\text{EnumDomain} & = (\lambda d : \text{Domain} \bullet \text{FreeTypeASF}(d.\text{name}, d.\text{literals})) \\
\end{align*}

Cards within the ATM system have an enumerated status variable representing the current state of the Card.

\begin{align*}
\text{CARDSTATUS} ::= \\
\text{ISSUED} | \text{INUSE} | \text{PINVALIDATED} | \text{REVOKED} \\
G\text{Domain} : \text{Domain} & \leftrightarrow \text{TYPE} \\
G\text{Domain} & = G\text{PlainDomain} \cup G\text{EnumDomain} \\
\end{align*}

### 2.4.2 Attributes

Attributes are defined as components of a schema for an entity. Where they are used in specifying the information that a process requires they will be declared as components of an Entity Model View Schema.

\begin{align*}
G\text{PlainAttrSigel} & \quad \text{Attributes are declared as declarations in the signature of a schema.} \\
G\text{PlainAttrSigel} : \text{AttrInst} & \leftrightarrow \text{SIGEL} \\
\text{dom} & \in G\text{PlainAttrSigel} = \{\text{ai} : \text{AttrInst} \mid \text{ai.card} = \text{ONE}\} \\
G\text{PlainAttrSigel} & = (\lambda \text{ai} : \text{AttrInst} \bullet \text{BscDeclASF}([\{\text{ai.name}\}, \text{TypeASF}(G\text{Domain}(\text{ai.domain}))])) \\
\end{align*}

Assuming entity Account that has an attribute accountNo.

\begin{align*}
\text{accountNo} : \text{ACCOUNTNO} \\
\end{align*}
Generating Formal Specifications from CASE Repositories

**GSetAttrSigel**  where the attribute has a cardinality of many it will be declared as set valued variable.

\[
\begin{align*}
\text{dom } GSetAttrSigel & = \{ ai : \text{AttrInst} \mid ai.\text{card} = \text{MANY} \} \\
GSetAttrSigel & = (\lambda ai : \text{AttrInst} \cdot \\
& \quad \text{BscDeclASF}(\{ ai.\text{name} \}, \text{TypeASF}(\text{FinsetASF}(\text{GDomain}(ai.\text{domain})))))
\end{align*}
\]

\[
\text{correspondenceAddress} : F \text{ ADDRESSLINE}
\]

\[
\begin{align*}
GAttrSigel & : \text{AttrInst} \mapsto \text{SIGEL} \\
GAttrSigel & = GPlainAttrSigel \cup GSetAttrSigel
\end{align*}
\]

\[
GAttrInstSigel : F \text{ AttrInst} \mapsto F \text{ SIGEL}
\]

\[
GAttrInstSigel & = (\lambda ais : F \text{ AttrInst} \cdot \{ a : ais \cdot GAttrSigel(a) \})
\]

2.4.3 **Entities**

For the plain entities a schema is declared using the entity name as the schema name. Each of the attributes is declared as a variable in the signature of the schema. Where an entity is a subtype it will include the schema of its parent in its signature.

**GPlainEntSch**  generates a schema for a plain entity. The entity name is used as the schema name and each attribute of the entity is declared as a component of the schema.

\[
\begin{align*}
\text{dom } GPlainEntSch & = \{ e : \text{rep.entities} \mid e.\text{type} = \text{PLAIN} \land e.\text{attributes} \neq \emptyset \} \\
\forall e : \text{dom GPlainEntSch} \exists s : \text{SCHEMA} \cup GPlainEntSch(e) = s \land \\
& \quad s.\text{name} = \text{SchNameASF}(e.\text{name}) \land \\
& \quad s.\text{sig} = GAttrInstSigel(e.\text{attributes}) \land s.\text{pred} = \text{TRUE}
\end{align*}
\]

Assuming the entity Account which has three attributes accountNo, correspondenceAddress and balance.

\[
\begin{align*}
\text{Account} \\
\text{accountNo} : \text{ACCOUNTNO} \\
\text{correspondenceAddress} : F \text{ ADDRESSLINE} \\
\text{balance} : \text{MONEY}
\end{align*}
\]

**GAssocEntSig**  where an entity is an associative entity it may not have any attributes. In this case the identifying attributes of the entities that it associates will be declared as components of the schema signature. If the associative entity has attributes these will be included as for a plain entity.
Generating Formal Specifications from CASE Repositories

\[ \text{GAssocEntSig : Entity} \rightarrow \text{SIGNATURE} \]

\[ \text{dom GAssocEntSig} = \{ e : \text{rep. entities} \mid \text{e.type} = \text{ASSOCIATIVE} \} \]

\[ \forall e : \text{dom GAssocEntSig} \cdot \exists \text{sig : SIGNATURE} \cdot \]

\[ \text{sig} = \text{GAssocEntSig}(e) \land \]

\[ \text{sig} = \{ \text{ai : AttrInst; s : SIGEL; r : rep. relationships} \mid \]

\[ e = r.\text{source} \land \]

\[ \text{ai} \in r.\text{dest.identifiers} \land \]

\[ s = \text{GPlainAttrSigel(ai)} \land s \} \]

\[ \lor \}

\[ \{ \text{ai : AttrInst; s : SIGEL} \mid \]

\[ e.\text{attributes} \neq \emptyset \land \]

\[ s = \text{GPlainAttrSigel(ai)} \land s \} \]

Assuming the associative entity Holding that has an attribute signatory to indicate whether the Holding has authorization for the account. The Holding also has the identifying attributes of the customer and account that it associates.

\[ \text{signatory : BOOL} \]

\[ \text{customerNo : CUSTOMERNO} \]

\[ \text{accountNo : ACCOUNTNO} \]

\[ \text{GAssocEntSch} \] the schema for an associative entity will have the signature defined by \( \text{GAssocEntSig} \) and an empty predicate.

\[ \text{GAssocEntSch : Entity} \rightarrow \text{SCHEMA} \]

\[ \text{dom GAssocEntSig} = \{ e : \text{rep. entities} \mid \text{e.type} = \text{ASSOCIATIVE} \} \]

\[ \forall e : \text{dom GAssocEntSch} \cdot \exists s : \text{SCHEMA} \cdot \text{GAssocEntSch}(e) = s \land \]

\[ s.\text{name} = \text{SchNameASF}(e.\text{name}) \land \]

\[ s.\text{sig} = \text{GAssocEntSig}(e) \land \]

\[ s.\text{pred} = \text{TRUE} \]

Assuming associative entity Holding.

\[ \text{Holding} \]

\[ \text{signatory : BOOL} \]

\[ \text{customerNo : CUSTOMERNO} \]

\[ \text{accountNo : ACCOUNTNO} \]

\[ \text{GSupTypeSigEl} \] where an entity is a subtype it inherits all of the attributes of its supertype. This is modelled by including the schema for the parent in the signature of the subtype.

\[ \text{GSupTypeSigEl : Entity} \rightarrow \text{SIGEL} \]

\[ \text{dom GSupTypeSigEl} = \{ e : \text{rep. entities} \mid \text{e.type} = \text{SUBTYPE} \} \]

\[ \forall e : \text{rep. entities} \cdot \]

\[ \text{GSupTypeSigEl}(e) = \text{InclASF(SchNameASF(e.supertype))} \]

\[ \text{GSubTypeSch} \] generates a schema for a subtype entity. The entity name is used as the schema name and each attribute of the entity is declared as a component of the schema. This schema only introduces the subtype entity and does not define the relationship between subtypes and supertypes.
Generating Formal Specifications from CASE Repositories

Assuming entity Account and the subtype Current Account.

<table>
<thead>
<tr>
<th>GSubTypeSch : Entity $\rightarrow$ SCHEMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{dom} \ GSubTypeSch = { e : \text{rep. entities} \mid e.\text{type} = \text{SUBTYPE} } $</td>
</tr>
<tr>
<td>$\forall e : \text{dom} \ GSubTypeSch \bullet \exists s : \text{SCHEMA} \bullet GSubTypeSch(e) = s \land$</td>
</tr>
<tr>
<td>$s.\text{name} = \text{SchNameASF}(e.\text{name}) \land$</td>
</tr>
<tr>
<td>$s.\text{sig} = \text{GAttrSigEl}(e.\text{attributes}) \cup { \text{GSupTypeSigEl}(e) } \land$</td>
</tr>
<tr>
<td>$s.\text{pred} = \text{TRUE} $</td>
</tr>
</tbody>
</table>

3. An Analysis of a Generative Approach

The process of implementing a generative approach to methods integration has been described and examples from each of stage of the process have been given. The purpose of implementing a generative approach is to provide the practitioner with a system for automatically generating formal specifications from the repository information that is defined for a non-formal method.

As with any approach to integration there are positive and negative aspects that will have an impact on the usability of the approach. These aspects can grouped into the following areas: the expenditure of effort that is required to implement a generative approach, the formal specification generated and the limitations that the approach imposes on the generation process. These aspects will have an effect on the utility of the generative approach as a mechanism for introducing the use of formal description techniques into the software development process.

3.1 Expenditure of Effort

The implementation of the a generative approach required the formal specification of the four stages. This formal specification contains 145 formal paragraphs. The meta-model description uses 40 schemata and 15 axiomatic or abbreviation definitions to specify the information content of the IPSYS OBJECT IE repository. The mapping functions consists of 91 axiomatic definitions of partial functions. The complete specification of the four stages of the approach, not including the source code for the implementation of the translation functions, consists of an 80 page document. The time taken to produce the documentation and implementation was approximately 4 months.

For a method such as IE the amount of work that needs to undertaken to implement a generative approach may at first seem prohibitive. However, the need for a more rigorous approach to methodology development has been recognised by researchers [5, 16, 15, 17]. The approaches that have been taken are to use formal techniques to specify the information content of a method and its associated semantics.

A related issue is that of customisation of established methods to suit particular projects and in-house needs. Hardy et. al. [13] found that 88 % of departments surveyed customise their method to some degree. 62 % of the respondents based this customisation on previous experience of tailoring methods. This is a concern as most methods have been developed to be internally consistent and rigorous, though normally not through the use of mathematical techniques, and there is a possibility of losing this consistency and rigour.
Researchers have understood the need for mathematical rigour in the development of methods and that most uses of methods are in a customised form. This need for rigour can be satisfied by the stages of the generative approach as the process involves a formal description of the information content of the method and a suite of rules which specify a semantically correct model in the method.

The benefits that can be accrued from taking this view are that the meta-model of the non-formal tool undergoes a comprehensive and rigorous specification of its content and meaning. Also there is good potential for highlighting omissions and contradictions that appear in the meta-model and generating a comprehensive understanding of the non-formal tool’s representational capabilities and limitations.

Another important benefit that can be gained from this process is in relation to the well-formedness rules that are defined as part of the four stage approach. These rules specify valid specifications that can be the subject of transformation and can be used to uncover areas where the CASE tool implementing the method could be enhanced with extra enforcement and verification techniques.

An example of this is the use of Entity Relationship Diagrams (ERDs). In the IE tool it is possible for an entity to be a plain entity, an associative entity or an attributive entity and for an entity to be of a different type in different ERDs. This conflicts with the formal semantics as it not possible to decide which well-formedness rule should be applied to the entity. The well-formedness rule for a plain entity requires that the entity has at least one attribute in addition to those attributes which identify the entity. Conversely the well-formedness rule for associative entities does not require the entity to have any attributes as the entity is used to resolve a many-to-many relationship between two existing entities.

In defining sub/super-type hierarchies the IPSYS tool shows implementation bias by insisting that subtype entities have a relationship between themselves and their super-type. This relationship between the subtype and supertype is used as the basis for two relational tables, one which holds the attribute values for the common attributes of the sub- and supertype and, one that holds the additional attribute values of the subtype and a foreign key for the supertype table.

One other area of concern is in the use of Activity Hierarchy Diagrams (AHDs) and Data Flow Diagrams (DFDs). The IPSYS tools allows activities on AHDs to be subordinate to themselves and processes in DFDs may appear on the decomposition of their own DFD. This is not reported by any checks that the tool offers except for a generic facility which reports where an object is used on a diagram.

### 3.2 The Generated Formal Specification

A generative approach has been used to translate the information content of a CASE repository into a formal specification. The success of the approach will depend heavily on the fitness for purpose of the formal specification. A case study has been produced using the IPSYS IE tool to populate the repository and provide input to the translation process. The formal specification generated is given in [25].

The term fitness for purpose will have different meanings depending on the use of the formal specification i.e. requirements specification, detailed design or documentation. In this work the aim of the generative approach is to automatically produce a detailed requirements specification that captured all of the information in the repository. The specification produced is incomplete in terms of constraints that may be required additionally to the constraints necessary to conform to IE semantics and predicates which define the effect of activities and processes in the normal manner of pre- and post-conditions.
For the purposes of this analysis the following criteria are used:

- structural clarity,
- traceability of information,
- verbosity and complexity,
- the use of language facilities.

While these criteria may not cover all the aspects of fitness for purpose they have been highlighted as significant in theoretical and practical work on method integration [10, 24, 2].

3.2.1 Structural Clarity

The structural clarity of the generated specification will have a significant impact on the readability of the specification and the separation of the three aspects of a system: data, behaviour and process. In Figure 2 the high level structure of the system specification is shown. Each aspect of the system is represented by one or more definitions (represented as rectangular boxes). The data aspect is represented by four definitions; entity, entity hierarchy, relationship and data model. The behaviour aspect is represented an entity life cycle definition. The process aspect by an activity definition.

Each of these definitions consists of a number of schema which capture one facet of the aspect. Taking the entity hierarchy definition as an example. The definition consists of a number of Subtype schema, Restriction schema, Partitioning schema and Supertype schema. The subtype schema describes the semantics of subtyping in terms of subsets of the ranges of the entity identification function. This injective function maps the ID of an entity to its instance and in the IPSYS IE tool all entities in a hierarchy have the same ID. The restriction schema provides a mathematical function to convert a subtype to a supertype using the Z concept of binding formation. The partitioning schema describes how the supertype instance set is partitioned by the subtypes. The supertype schema describes the instances of supertypes in terms the instances sets of subtypes.
These four schema together with the mechanism of schema inclusion used in the entity type schema to inherit attributes from the supertype provide a clear way to model the four characteristics of subtype instance sets.

- coverage
- exclusivity
- orthogonality
- inheritance

The structure of the other definitions in terms of schemata also separates the characteristics of the objects under consideration.

An advantage of this separation also becomes apparent when the formal specification is extended with constraints. The separation allows the practitioner to concentrate on a single aspect of an object at one time and to record any constraints that may be relevant to that aspect while allowing constraints that apply at a higher level to be defined separately. It should be noted that there is an obligation to ensure that no contradiction is introduced through these separate constraints, though this is true of a hand written formal specification in any case.

3.2.2 Traceability

Much of this structure, though not all, is derived from the structure of meta-model components that the formal specification components relate to. This promotes good traceability between the non-formal and formal specifications that is considered vital when changes are required to be made to either the non-formal or formal. Where the change is made from the non-formal the situation is eased by the generation process that in effect produces a new formal specification. However, it is likely that once the incomplete formal specification has been developed with additional constraints and activity definitions there will be a need to revisit the non-formal specification to make changes. The close structural correspondence will aid this task in identifying where changes are required.

3.2.3 Verbosity and Complexity

One of the criticisms (see [2] page 36) made against formal specifications that closely follow the structure of the non-formal is:

If the form of the Z is kept close to the informal analysis, it tends to result in a verbose and cluttered specification that lacks the elegance that is essential for understanding

While this true in principal it does not always apply when the formal specification is used as a documentation tool that has formal requirements for formatting and structure. Also within the scope of the work reported this well defined structure and explicit style of specification is an advantage when considering practitioners who are not experts in writing formal specifications.

3.2.4 Use of Language Facilities

Hall [11] has identified the problem of following the structure of the non-formal specification too closely in that it ignores the richer and more expressive facilities of most formal languages as compared to graphical representations. The formal specification generated does not use all the language facilities to the best advantage, though schema inclusion is used widely to structure the specification. Further the specification can be improved through the use of promotion and the schema calculus when the activity definitions are fully specified. This lack of use of language facilities will be addressed in the improved generative approach of Section 4.
Generating Formal Specifications from CASE Repositories

The generated specification could be improved either by using a different representation in the formal language, as has been done subsequently by the author, or by choosing an alternative formal representation that is better suited to the task by having first objects and constructs which capture the semantics of techniques such as entity hierarchies and subtyping. However, the main objective of implementing a generative approach was to investigate the process itself and to attempt to improve and generalise the process while removing as many negative aspects as feasible.

3.3 Summary

Although the generative approach has been shown to be a viable way of producing a formal specification from the products of an non-formal method [1, 22, 14] there are important aspects that need to be considered when an implementation is to be attempted [27]:

- the approach requires an expenditure of effort on the four stages.
- the formal specifications produced can be verbose.
- certain meta-model constructs do not formalise well.
- the conceptual shift from the non-formal notation to the formal.

There are certain negative aspects of the generative approach that should be addressed to improve the approach:

- additional structuring mechanisms and conventions are required.
- there is loss of information in the generation process.
- the meta-model description is targeted to one particular methodology.
- the practitioner may want to hand-write parts of the formal text.

The next section will describe an improvement of the generative approach that addresses these negative aspects. The improvements to the approach are summarised below:

- the practitioner can chose when to use formal language directly and so not lose the expressive power of the language.
- any part of the non-formal specification not explicitly given a formal text can have a default representation.
- the structure of the non-formal specification is maintained.
- the generation process is generalised to a wider spectrum of non-formal methods.
- different formal representations of the meta-model constructs may be defined for use in different projects and within different stages within a project.

4 Improving the Generative Approach

In the generative approach the formal description of the meta-model is tied to one particular implementation of a methodology. An improvement of the generative approach would remove this bias towards a particular meta-model and allow the approach to be generalised to a wider spectrum of methods (see [18] Page 335).

The generative approach requires the formal specification of the input to the mapping functions. Any approach to removing the bias still requires the meta-model information to be well defined. An alternative means of representing meta-model information is needed that is not tied to one particular method and is capable of representing the information requirements of a wide spectrum of methods.
Generating Formal Specifications from CASE Repositories

The use of a representation that meets this need allows the techniques of a generative approach to be applied to more than one non-formal method without having to re-write the meta-model description for each method. The more general representation will be formally defined once and once only. This allows any vendor of a CASE tool for a non-formal method to export the semantic information from their tool’s repository to the unbiased representation. The formal specification can then be generated from this representation.

4.1 The CASE Data Interchange Format

The CASE Data Interchange Format (CDIF) [6] collection of standards is an approach to providing a comprehensive framework to address the needs of CASE tool users and vendors in transferring semantic information between CASE tools. The standards define a single interface to enable the import and export of information together with the semantics and presentation of this information. There are three major sections to the CDIF standards: the underlying architecture of CDIF standards and the mechanism for extending the meta-models; the transfer syntax and format which provides the import/export interface; and the collection of standards which form the Integrated Meta-model. The CDIF Integrated Meta-model is divided into sub-sections called subject areas. Each subject area addresses either a particular aspect of the development life cycle, such as entity-relationship modelling, or it addresses some aspect of commonality used by other subject areas.

These subject areas define meta-models that are capable of representing the semantic information for a wide variety of techniques and methods. These meta-models will be used to replace the method specific meta-models used in the original generative approach.

4.2 Using the Formal Notation Directly - The Incremental Adoption of Formal Specification Techniques

An improvement would allow the use of the formal specification notation directly by the practitioner. This would allow the richer and more expressive formal notation to be used where the software specification can benefit most from rigour and precision.

In a simple case this could be where the practitioner has written the formal specification of an object rather than use the default translation. In a more complex case the practitioner may have recognised that a group of entities and relationships could be better represented directly in the formal notation using the richer structures that are available e.g. sequences or bags.

If the practitioner is to hand-write fragments of the formal specification these must be included with the generated parts in an orderly manner. This will be further discussed in Section 4.3.3.

The application of a generative approach where the practitioner can incorporate custom written fragments can be viewed as a starting point at which practitioners adopt the use of formal specification techniques. The generation of the formal specification allows the practitioner to gradually take over the process of writing formal specifications though ensures that the structure of the specification reflects the structure of the informal specification. As the practitioner gains experience in the use of the formal notation the amount of the specification that is automatically generated should reduce.
It should be noted that an incremental approach must be supported by education, training and expert advice. The automatic generation can support the adoption of formal techniques by alleviating the burden of producing a complete specification and by imposing a good structure on the formal specification but it cannot replace the need for a thorough understanding of the problem domain and specification techniques.

4.3 Managing the Use of Hand Written Fragments

In the improved generative approach the CDIF Integrated Meta-model is used to capture the semantic information from the non-formal tools repository that is then used as the input to the generation of the formal specification. This allows for the automatic production of the formal specification as an initial approach. As the practitioner adopts the direct use of the formal specification language it is necessary to ensure that the traceability from non-formal to formal is maintained. The CDIF General Structuring Mechanism is used to provide a traceability mechanism and the inclusion of hand-written specification fragments.

4.3.1 The CDIF General Structuring Mechanism

The CDIF suite of standards use a mechanism to structure the decomposition of definitions into components that may be reused and shared in other definitions. This mechanism, the General Structuring Mechanism (GSM), uses a number of CDIF meta-entities and meta-relationships to achieve this structuring. Figure 3 gives the meta-model for the objects used. The two major meta-entities are the ComponentObject and the DefinitionObject.
The ComponentObject uses a meta-relationship References to link components to their definitions. The DefinitionObject has two meta-attributes, SpecificationLanguage and SpecificationText, which capture the specification of the DefinitionObject and define the language used for the specification. The DefinitionObject also has a meta-relationship Contains that links the object to its structural decomposition of components. The ComponentObject has two subtypes ReferencedElement and EquivalenceSet that provide a mechanism for accessing distinct instances of components that are shared in definitions and allow for identical instances of components to be identified.

4.3.2 Use of the GSM

The GSM is used to provide a facility where formal specification fragments are linked to the non-formal object that they represent. To achieve this four additional meta-entities and three meta-relationships are defined using the CDIF extensibility mechanism, see Figure 4. The meta-entity ComponentObject is extended with three subtypes, Structure, Definition and Constraint.

The Structure meta-entity is used to represent structural components. The Definition meta-entity is used to represent definitions of structural components. The Constraint meta-entity is used to represent constraints on structural components and their definitions. The Constraint meta-entity can be linked to a Structure or Definition meta-entity using the meta-relationship Constrains.

The third meta-entity SpecificationObject is a subtype of DefinitionObject that is used to define the specification of the object in the formal language using the meta-attributes SpecificationLanguage and SpecificationText inherited from DefinitionObject. The meta-entity SemanticInformationObject, which is the supertype of all CDIF meta-entities that carry semantic information, is extended with two new meta-relationships IsFullySpecifiedBy and IsPartiallySpecifiedBy. This allows any object that carries semantic information to be linked to the SpecificationObject that fully/partially defines its specification in the formal notation. Where an object has only been partially specified in the formal notation the default translation will still be performed on those components that are defined in the CDIF transfer.

Figure 4: Using the General Structuring Mechanism

The GSM is used to provide a facility where formal specification fragments are linked to the non-formal object that they represent. To achieve this four additional meta-entities and three meta-relationships are defined using the CDIF extensibility mechanism, see Figure 4. The meta-entity ComponentObject is extended with three subtypes, Structure, Definition and Constraint.

The Structure meta-entity is used to represent structural components. The Definition meta-entity is used to represent definitions of structural components. The Constraint meta-entity is used to represent constraints on structural components and their definitions. The Constraint meta-entity can be linked to a Structure or Definition meta-entity using the meta-relationship Constrains.

The third meta-entity SpecificationObject is a subtype of DefinitionObject that is used to define the specification of the object in the formal language using the meta-attributes SpecificationLanguage and SpecificationText inherited from DefinitionObject. The meta-entity SemanticInformationObject, which is the supertype of all CDIF meta-entities that carry semantic information, is extended with two new meta-relationships IsFullySpecifiedBy and IsPartiallySpecifiedBy. This allows any object that carries semantic information to be linked to the SpecificationObject that fully/partially defines its specification in the formal notation. Where an object has only been partially specified in the formal notation the default translation will still be performed on those components that are defined in the CDIF transfer.
It should be noted that the extensions to the GSM are only used when the formal specification of an object is decomposed into structural components. This may result from reuse of formal specification fragments. Where an object’s formal specification is not decomposed it can be represented directly in the \textit{SpecificationObject} which is linked to it via one of the meta-relationships \textit{IsFullySpecifiedBy} or \textit{IsPartiallySpecifiedBy}.

4.3.3 \textit{Using the Extended GSM}

An example from the ATM subsystem will be used to demonstrate the use of the extended GSM.

In the ATM subsystem example the accounts that a customer holds are represented by three entities; Customer, Account and Holding and two relationships; OwnsAndOwnedBy and ParticipatesInAndHeldBy. Where a customer is a signatory for an account this is recorded as an attribute of the Holding entity. We may choose to use the formal specification language directly by rewriting the definition of customer holdings. This new definition consists of two new attributes and a formal definition of a Boolean function which returns True if a Customer is a signatory for an Account and False otherwise.

Figure 5 shows the structure of the instances in the CDIF transfer. The formal specification generated from this CDIF transfer is given next.

\textbf{Generated Z Specification}

\begin{verbatim}
Customer
customerNo : CUSTOMERNO
name : NAME
address : TEXT
holdings : F Account

| customers : F Customer
\end{verbatim}
Generating Formal Specifications from CASE Repositories

accountNo : ACCOUNTNO
correspondenceAddress : F ADDRESS
balance : MONEY
signatories : F NAME

accounts : F Account

signatory : Customer × Account → Boolean

∀ c : Customer, a : Account | (c, a) ∈ dom signatory •
signatory = (λ c : Customer, a : Account •
  if c.name ∈ a.signatories then True else False)

dom signatory = { c : customers; a : accounts }

The generated formal specification includes the hand written fragments within the schemata for Customer and Account. The definition of the function signatory is presented as an axiomatic definition. If the specification object is named then it will be presented as a schema. Other forms of formal representation can be defined by extending the specification object with an attribute describing the type of specification object or by subtype classification.

5 Summary and Conclusions

A method for formally specifying the generation of formal specifications from a CASE tools’ repository, as defined by a meta-model has been described. Although this approach has been shown to be effective there are a number of limitations that this approach imposes in terms of the effort required in implementing the approach and the restrictions imposed by the inflexibility of the approach.

An alternative approach has been outlined that addresses an important limitation, that of the original approach being method specific. This allows the approach to be generalised to other methods through the use of a neutral representation and by viewing the generation of a formal specification as a transfer between two CASE environments. The neutral representation, the CDIF Integrated Meta-model, is a published suite of standards that defines an integrated meta-model that addresses the information requirements of various aspects and stages of the software development process. The standards also define an import/export format for transferring information between CASE implementations. The CDIF standards chosen to generalise the generative approach also provides a mechanism for including custom written formal specification fragments in a controlled way.

The work addresses an important aspect of the use of formal specification techniques, it’s lack of wide penetration, in the general software development community by introducing techniques that allow their incremental adoption by practitioners.

6 Future Work

The approach will need to be extended to cover the other Subject Areas that have been published or are in production. The support for different specification languages and constructs should be improved by defining new meta-entities and meta-relationships to represent these languages and constructs. The software support that has currently been developed for the improved generative approach will need to be extended to other CASE implementations and platforms.
Each of the items of interest has a name:

```
  [NAME].
```

Decorations may be applied to schema names.

```
  DECORATION ::= dash | Jdot | out
```

Where names are used as schema names there are functions defined that translate names into schema references.

```
  SCHEMA REF ::= SchNameASF | NAME | SchDecorASF | NAME | DeltaASF | NAME | XiASF | NAME | HidingASF | NAME | seq | NAME
```

A number of expressions are used.

```
  EXPR ::= DomASF | InvASF | OccASF | SelASF | AplASF | SetExpr | MapExpr | CompExpr | SchRefExpr | SchRef | TypeAST | LamExpr | MuExpr | ThetaExpr | RanASF | IntersectAST | UnionAST | EmptySet
```

Predicates are defined as sets of predicate elements.

```
  PREDICATE ::= ElemSet | Conjoin | Disjoin | TRUE | FALSE
```

A schema has a signature that can contain a set of signature elements.

```
  SIGNATURE ::= F SIGEL
```

```
  name : SCHEMA REF
  sig : SIGNATURE
  pred : PREDICATE
```
8 References


Generating Formal Specifications from CASE Repositories


