

On the Specification of a Component Repository

Nuno Rodrigues¹ and Luis Barbosa²

¹ Sidereus, S. A., Porto, Portugal
nfr@sidereus.pt

² Dep. Informática, Univ. Minho, Braga, Portugal
lsb@di.uminho.pt

Abstract. The lack of a commonly accepted definition of a software component, the proliferation of competing ‘standards’ and component frameworks, is here to stay, raising the fundamental question in component-based development of how to cope in practice with heterogeneity. This paper reports on the design of a Component Repository aimed to give at least a partial answer to the above question. The repository was fully specified in VDM and a working prototype is currently being used in an industrial environment.

1 Introduction

“I’m sure this problem has already been solved” and “probably, others would like to (re)use my solution”, are common concerns in the everyday life of a programmer. Concerns that the emerging of *component-based* programming paradigm aims to transform into effective programming practises. Actually, all engineering disciplines rely on standard *components* to design and build their artifacts and development methodologies based on *third-party assembly* of components — software engineering should not be the exception.

Although the term *software component* has been around for a long time, *component-based programming* has become a buzzword since mid 1990’s (see, e.g., [7, 11, 10, 6, 12]). The basic motivation is to replace conventional programming by system’s construction by composition and configuration of reusable off-the-shelf units, often regarded as ‘*abstractions with plugs*’. Or, quoting from [11], ‘*independently deployable unit of composition with contractually defined interfaces*’. The paradigm is often illustrated by the visual metaphor of a *palette* of computational units, treated as black boxes, and a *canvas* into which they can be dropped and plugged together through *wires*, corresponding to some sort of gluing code.

However, as it happened before with *object* orientation, component programming has grown up to popular technologies before consensual definitions and principles, let alone formal foundations, have been put forward. In particular, it has given rise to a number of increasingly popular technologies designed around specific *interaction paradigms* — e.g., CORBA and JAVABEANS based on object-like composition, JAVASPACEs and TSPACEs on top of *shared data space* models, typical of classical coordination languages, and *channel* based techniques

inherited from process languages, as in IWIM/MANIFOLD or DRAWIN. Such technologies also differ at the application level — *e.g.* acting as source-level languages extensions (such as CORBA or JAVASPACEs), or as connector frameworks for dynamic (binary-level) units (as in *e.g.*, COM and .NET).

From an engineering point of view a key issue to the success of the component paradigm is *integration*. In fact, unless one restricts himself to a particular framework or, even worse, to a specific tool within a framework, it is no easy task to identify, select, re-use and compose heterogeneous software components from a virtual ‘global market’ (as represented, for example, by the Internet). The lack of a commonly accepted definition of ‘component’, the proliferation of competing ‘standards’ and component frameworks, is here to stay, raising the question of how to cope in practice with heterogeneity.

This paper reports on a concrete approach to a component integration problem found in the context of an european-wide industrial software development project¹. The challenge was to build a component repository (of both source code and interface information) able to register and marketing all the components produced by different teams in the project. The repository was also supposed to act as an *exchange market*: each user being able not only to register its own components, but also to announce its plans to issue specific components and even to publicise its needs and associated requirements in the form of collections of APIs. Although such *component marketplaces* are emerging in the Web, the underlying description techniques remain rather informal, mainly textual, plus an additional, limited, classification in terms the intended business areas.

In a sense API, the acronym for *Application Program Interface*, emerged as a keyword. But what is in an API, when the component technology used ranges from JAVA and CORBA code to HASKELL or WSDL? Coming back to Szyperski’s definition above, what seems crucial to handle is the definition of what an *interface* is.

Initial prospects of implementing the repository as an API database and resorting to text-based retrieve methods for querying, soon reveal impractical given the heterogeneity of the APIs supplied by the the different project teams. As the complexity of the individual components and the size of the repository increases, higher levels of abstraction are required, and, for all practical purposes, going abstract means going formal. The VDM meta-language [4, 3] was chosen as the common language to which different sorts of APIs were mapped. A complete, working prototype of the repository, named COMPSRC to express the idea of a ‘component source indexing space’, has been developed in the VDM TOOLBOX complemented with a dedicated web interface for easier access.

The paper is organised as follows: the COMPSRC architecture and the generic component model underlying the whole system are described in the following two sections. Section 4 discusses some of the composition patterns automatically detected by COMPSRC. We conclude with some prospect of future work in section 5.

¹ The Eureka IKF initiative.

2 The COMPSRC Architecture

The COMPSRC repository is built around a broad characterisation of what an interface to component source code is, specified in the VDM meta-language in the early project stages. This is called, in the context of this system, the *abstract API*. The component source code database is indexed by such abstract APIs which are ‘extracted’ from the actual component APIs. This is done by extractor functions specified for the typical target languages used in the project. Such extractors correspond to the incoming arrows in Figure 1.

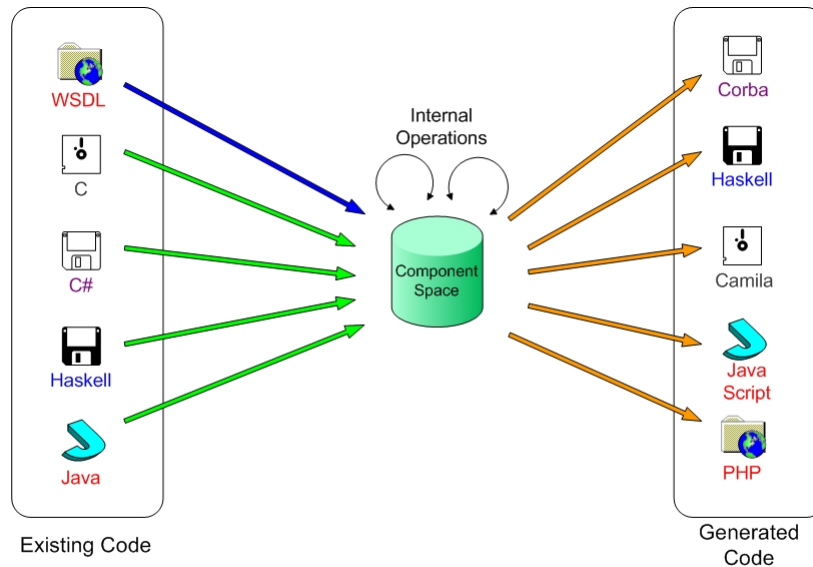
In the centre of the structure diagram is the interface repository indexing the actual component source data base. The repository supports the usual management operations — *e.g.*, registering, removing, updating, etc. The most relevant feature, however, is the support of a *component calculator* which makes it possible to compute new services based on the automatic composition of old functionality, according to more or less strict interaction patterns, and, then, to build new APIs and new components which would act as if the actual source code has been merged. The interaction patterns available are different kinds of *pipelining* (linear, flattened or monadic, as described below) as well as *multiplicative* and *additive* aggregation, corresponding to component parallel composition and choice, respectively. The underlying component calculus is described in detail in [2, 1]. Once a new component is built, by the calculator, the new abstract API is registered in the system and the component ‘combined’ code is supplied, in the form of a conditional compiling script in the .NET notation. This corresponds to the outgoing arrows in the diagram.

3 Abstract Interfaces

The ‘quest’ for a common *abstract API* format started from a reverse specification of the (static information part of) different sorts of component APIs used in the project. These ranged from the ‘object-oriented flavour’, emphasising hierarchical class structures, to ‘functional modules’ described by service signatures and, eventually, datatype constraints. The formal analysis of typical API examples in JAVA, CORBA, MICROSOFT VISUAL BASIC, HASKELL and WSDL, which is documented in [8], lead to the description of a component as an indexed collection of services (referred to as *modules* in the sequel) and mirroring the overall component architecture described in Figure 2. Formally,

Each module is, then, defined as an ordered tuple of information including the following aspects:

1. **Desc** - a text description of the current module
2. **INH** - a set of modules imported by this module
3. **Ex** - a map enumerating examples concerning the module usage
4. **Func** - the set of functions (fine-grain services) defined in the module
5. **State** - the implementation state of the module which is one of :
 - **Implemented** - the module is implemented and ready to use

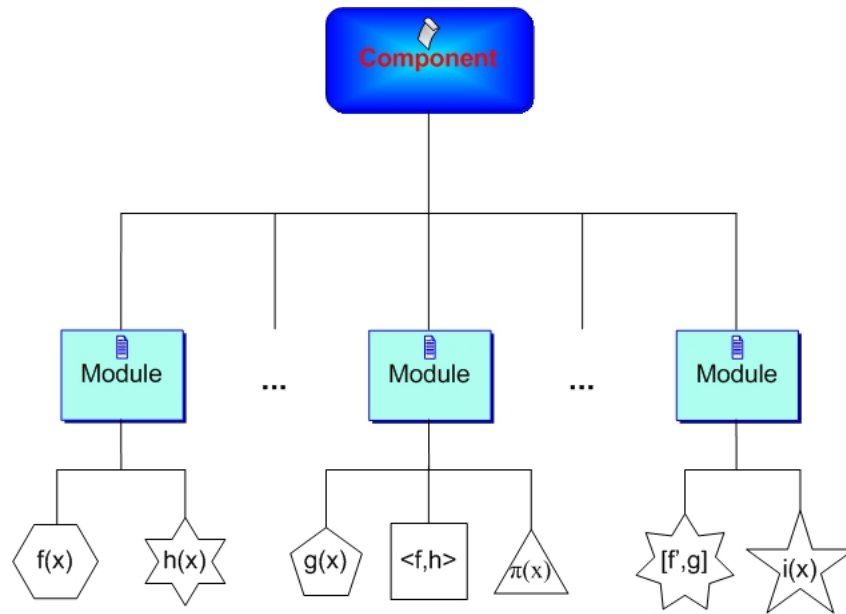


1. System Architecture

- **In Implementation** - the module is currently in implementation process
 - **Requested** - someone is interested in this module and is willing to pay for it
 - **To Convert** - the module is specified in a language but needs to be implemented in a different platform
 - **Generated** - generated by the 'component calculator' in COMPSRC
6. **Platform** - the target implementation platform
 7. **StatDep** - the static dependencies set, i.e., environment variables, runtime environments, ...
 8. **Protocols** - a brief description of the communication protocols that the module implements
 9. **SpecAPI** - specific (platform-dependent) module data for each kind of API, basically distinguishing purely functional interfaces from state-based ones (as found in object-orientation)

4 Component Assembly

For each API submitted to COMPSRC a corresponding *abstract API* (i.e., a value of the respective VDM datatype) is built by the (language-dependent) *interface extractors* mentioned above. Such abstract APIs provide the basis for classifying, locating and retrieving components from the underlying repository. Furthermore



2. Component Architecture

```

CompData :: Modules : ModuleId  $\xrightarrow{m}$  ModuleData
          Name : String
ModuleId =  $\mathbb{Z}$ ;

```

they become the ‘raw material’ used by the *component calculator* to generate new components.

Such generation is done in two different, but complementary, ways: *aggregation* and *wiring*.

Component *aggregation* is achieved by the application of a small set of operators acting on the abstract interfaces as a whole ². In particular, they model:

- Interface *restriction* to a set of *modules* or even, within a particular module, to a set of services.
- *Additive aggregation*, in which different modules of two different interfaces are selected and packaged into a new one. If both interfaces are indexing actual code written in compatible implementation languages (*i.e.*, related by an *embedding*) the code generation process is activated.
- *Multiplicative aggregation*, corresponding to the synchronous execution of modules in two different components.

² see [2] for an overview of the corresponding calculus partially implemented in COMP-SRC.

```

ModuleData :: Desc : [token]
            INH : Inheritance-set
            EX : Examples
            Funcs : Func-set
            State : IMPLEMENTED | INIMPLEMENTATION |
                  PAYTOIMPLEMENT | MODULED |
                  GENERATED
            Platform : [String]
            StatDep : [token]
            Protocls : token-set
            SpecAPI : SpecOOAPI | SpecFAPI
Examples = ExId  $\xrightarrow{m}$  token;
ExId =  $\mathbb{Z}$ ;

```

The *wiring* process, on the other hand, consists in the search for *composition possibilities* among the module collections of a specified set of components. Such search can be systematic, exposing all possible connections arising from a given set of components, or user-oriented in which each possible composition is validated or discarded by the user.

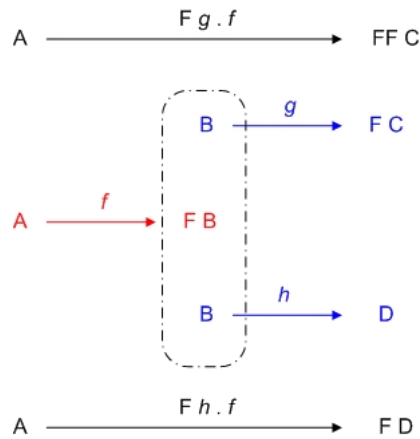
Recall, from the module specification above, that its kernel is a collection of functions modelling the fine-grain services made available. In general each such function is specified as

$$FI \rightarrow GO$$

where I and O are the *import* (respectively, *export*) datatypes embedded in a behavioural context F (respectively, G) represented as a polynomial functor. For example a non deterministic service may be modelled by a function $\text{serv} : I \rightarrow \mathcal{P}O$, where \mathcal{P} stands for the (finite) powerset functor. Similarly a service in the form of a partial transducer may take the form of $\text{machine} : I \rightarrow \mathbf{1} + (O \times I)^K$.

This sort of embedding through a functor entails the need to equip COMPSRC with the ability to perform different types of ‘functional’ composition, besides the obvious one between functions sharing the domain of one with the codomain of the other. In particular the following ‘extra’ *pipelining* composition patterns are considered:

- *Curry insensitive*, which allows to perform curry or uncurry on a pair of otherwise no composable services.
- *Monadic*, whenever the context information is captured by a *monad* (which, as shown in [1] is often the case). Monadic composition includes both the usual Kleisli composition of monadic functions used in functional programming, and a simpler scheme amounting to the monadic embedding of a ‘plain’ function. Both cases are illustrated in Figure 4.



4. Monadic Interaction

The specification of the wiring process is based on a structural comparison function to relate, up to the intended composition pattern, functions input/output contexts. Function *equalsLessFunctor* below, defined by induction on the structure of polynomial functors, sets the underlying datatype comparison.

5 Conclusions and Further Work

Software development by component assembly is most likely to become the main stream in software engineering in the near future. This will lead to a broader understanding of what a software component is (virtually any sort of content can be encapsulated in a reusable entity with well-defined interfaces and able to be connected at runtime) and to the emergence of standardised component frameworks able to integrate heterogeneous components and deal with other such systems in a cooperative manner. The ability to deal with ‘non native’ components has been recognised as the hallmark of the so-called ‘second-generation’ component systems [9], a step ahead of what is currently achieved in CORBA, JAVABEANS or OBJECTSPACE VOYAGER.

The basic lesson learnt from the development of COMPSRC is the potential of formal, model-oriented, methods in guiding the design of such platforms. We believe this exercise can be further extended to cope with some issues not covered in the present version, namely, dynamic instantiation of components, location and mobility. Those are fundamental issues for modelling *distributed* component frameworks.

The COMPSRC prototype has been used, not only in the context of the project in which it was originally developed, but also to organise software components

```

equalsLessFunctor : DataType × DataType →  $\mathbb{B}$ 
equalsLessFunctor (fdt, dt)  $\triangleq$ 
  cases fdt :
    mk-Plus (s) → dor ([dor ([compareFunctor (dt.P (i).#2, s (j).#2) | i ∈
inds dt.P]) | j ∈ inds s]),
    mk-Times (s) →
      is-Times (dt) ∧
      len s = len dt.P ∧
      dand ([compareFunctor (s (i), dt.T (i)) | i ∈ inds s]),
    mk-Power (b, e) → compareFunctor (b, dt.Base) ∨
      compareFunctor (e, dt.Exponent),
    mk-Map (l, r) → compareFunctor (l, dt.Left) ∨
      compareFunctor (r, dt.Right),
    mk-Set (d) → compareFunctor (d, dt.Of),
    mk-Seq (d) → compareFunctor (d, dt.Of),
    mk-Curry (-) → false,
    mk-Var (-, t) → is-Var (dt) ∧
      dt.Type = t
end;

```

arising from a massive re-engineering effort of legacy code undertaken by the software company to which the first author is affiliated.

Such an architectural re-engineering effort aimed to identify service components orthogonal to the basic development layers considered in the design practice (*i.e.*, database definition, middleware and GUI). For each identified component an abstract interface, as described above, has been written and directly submitted to COMPSRC together with a .NET script to navigate in the (monolithic) legacy code (which remains unchanged) and generate the actual executable code corresponding to the new abstract API. In a subsequent stage new software products incorporating the recovered components have been generated within COMPSRC.

Future work on the the COMPSRC prototype is foreseen in two main directions:

- The scaling up COMPSRC to act over the web in a *transparent* way, instead of relying on a localised component source database, as well as making it able to produce new component connection at runtime, is a main challenge from a technological point of view.
- Conceptually more demanding is the addition of facilities to cope with heterogeneity not only at the ‘linguistic’ and component-style levels (*e.g.*, the integration of object and functional models), but also at the level of the *interaction style* (aiming at the integration of, *e.g.*, method invocation, dataflow stream processing and event-based interaction). Some preliminary work on a similar topic is documented in [5].

References

1. L. S. Barbosa. Towards a calculus of state-based software components. In *Proc. of 7th Brazilian Symposium on Programming Languages (selected papers)*, Ouro Preto, Brasil, June 2003.
2. L. S. Barbosa and J. N. Oliveira. State-based components made generic. In H. Peter Gumm, editor, *Elect. Notes in Theor. Comp. Sci. (CMCS'03 - Workshop on Coalgebraic Methods in Computer Science)*, volume 82.1, Warsaw, April 2003.
3. J. Fitzgerald and P. G. Larsen. *Modelling Systems: Practical Tools and Techniques in Software Development*. Cambridge University Press, 1998.
4. Cliff B. Jones. *Systematic Software Development Using VDM*. Series in Computer Science. Prentice-Hall International, 1986.
5. K.-P. Lohr. Towards automatic mediation between heterogeneous software components. ENTCS, volume 65.4, Elsevier, 2002.
6. B. Meyer and C. Mingins. Component-based development: From buzz to spark. *IEEE Computer*, 32(7):35–37, 1999.
7. O. Nierstrasz and L. Dami. Component-oriented software technology. In O. Nierstrasz and D. Tsichritzis, editors, *Object-Oriented Software Composition*, pages 3–28. Prentice-Hall International, 1995.
8. N. Rodrigues. Abstract interfaces and extractores. Technical report, Sidereus and DI (U. Minho), 2003.
9. K. Schmaranz. On second generation distributed component systems. *Journal of Universal Computer Science*, 8(1):97–116, January 2002.
10. J.-G. Schneider and O. Nierstrasz. Components, scripts, glue. In L. Barroca, J. Hall, and P. Hall, editors, *Software Architectures - Advances and Applications*, pages 13–25. Springer-Verlag, 1999.
11. C. Szyperski. *Component Software, Beyond Object-Oriented Programming*. Addison-Wesley, 1998.
12. P. Wadler and K. Weihe. Component-based programming under different paradigms. Technical report, Report on the Dagstuhl Seminar 99081, February 1999.