Meta-Data Based Design of Workflow Systems

A. Barry[1], N. Baker[1], J.-M. Le Goff[3], R. McClatchey[1], J.-P.Vialle[2]

[1]*Centre for Complex Cooperative Systems, Univ. of the West of England, Bristol BS16 1QY, UK*
email: Richard.McClatchey@uwe.ac.uk

[2] *LAPP, IN2P3,Annecy-le-Vieux, France*
email:Vialle@lapp.in2p3.fr

[3] *EP Division, CERN,1211-Geneva 23,Switzerland*
email:Jean-Marie.Le.Goff@cern.ch

Abstract

The next generation of workflow systems will need to interoperate with non-workflow systems, to handle distribution of workflows and to cope with dynamically changing workflow definitions and ad-hoc workflows. Current commercial systems do not cater for these aspects; any new system must provide these facilities from the bottom up. Rather than design a workflow system to cope with these second generation workflow requirements by generalising from a specific model of workflows, this position paper advocates the adoption of a meta-data approach to developing a workflow object model. An example is given from the CRISTAL project where a meta-model has been designed based on generic definition objects.

1. Introduction

Up to now, there have been relatively few examples of the application of workflow management outside the business domain. Workflow management allows the combination of a data-oriented view on applications, which is the traditional one for information systems, with a process-oriented one in which activities and their occurrences over time are modelled and supported properly. Since workflow management combines influences from a variety of disciplines, including cooperative information systems, computer-supported cooperative work, groupware systems, or active databases, it has recently attracted the attention of non-business application domains. Two of these, the domain of scientific applications (in particular in the natural sciences) and that of engineering applications, seem particularly appropriate for the exploitation of workflow technology, since they involve processes in which humans and machines interact in considerable numbers, and could benefit from the automation in the execution of such processes. However, the requirement for workflow management in these areas differs significantly from those in business. Consequently, there has been little work so far towards an in-depth understanding of the relevant issues.

During the last few years, workflow management has become increasingly interesting to scientists and to engineers who, for example, carry out experiments in a laboratory, design work in a computerised environment or perform once-off apparatus construction. What such applications have in common is the fact that the processes to be executed are frequently (sequences of) events with outcomes which can evolve as the experiment advances, so that the structure of the entire process is difficult to determine in advance. Nevertheless, modelling, execution control, and documentation (for the purpose of reuse) are highly relevant.

From a workflow point of view, the important features found in these applications include:

• flexibility in structuring and modelling (open-ended, sometimes ad-hoc workflow definition, allowing decision-making whilst a workflow is being executed);
• workflows with a complex (or nested) inner structure of individual steps (such that multi-level modelling becomes appropriate);
• distribution of workflow execution;
• the treatment of failures which can be more complex than dealing with ordinary cases;
• system functionality features such as browsing and visualization, documentation, or coupling with external tools, e.g., for analysis.

Moreover, their workflow execution requirements ask for features like:

- support for long-running activities with or without user interaction
- application-dependent correctness criteria for executions of individual and concurrent workflows
- integration with other systems (e.g., file managers, DBMSs, Product Data Managers) that have their own execution/correctness requirements
- reliability and recoverability with respect to data
- reliable communication between workflow components and processing entities.

The design of these second generation workflow systems (elsewhere referred to as *scientific workflows*) requires that the model used for definition and execution be sufficiently flexible to cater for these needs. It is the view of the authors that an object-oriented design approach based on a philosophy of *meta-objects* satisfies these requirements.

This position paper outlines this philosophy in the context of the CRISTAL project being developed at CERN to support the construction of large-scale engineering apparatus. In CRISTAL a UML [1] meta-object model has been used to provide aspects of interoperability between workflow and non-workflow systems [2] and to provide reflective [3] aspects to ease model navigation. The first section details design constraints in CRISTAL and proposes a meta-modelling approach. The following section shows how this approach has lead to ease in integrating workflow and PDM systems and the final section demonstrates how the CRISTAL meta-model supports multiple views onto the CRISTAL repository.

2. Design Constraints in CRISTAL

The CERN CRISTAL project [4, 5] combines many of the requirements of second generation workflow systems. This section uses the CRISTAL project as a vehicle in which to demonstrate aspects of design constraints for scientific workflows. Firstly, the environment at CERN is research-based and both workflow and product-related definitions tend to evolve rapidly over time. The CRISTAL software must cater for the development of a High Energy physics detector (CMS [6]) which will take place over an extended period of time (1999-2005) and whose design will naturally advance as time elapses. As a consequence of this CRISTAL must also support long-running and potentially nested workflow activities, with natural consequences on transaction handling.

Secondly, the construction of CMS is a once-off process. In other words, the evolution of workflows and product data must be allowed to take place as the production continues - versions of workflow activities and product definitions must co-exist in the production process for the duration of CMS construction. In addition, users of CRISTAL must ultimately be able to cater for the ad-hoc definition and execution of workflows, as and when required in CMS construction.

Thirdly, the CMS construction process is highly distributed. Production of (versions of) CMS products will take place in areas as disparate as China and Russia, their testing will be undertaken in Rome, CERN and the UK and assembly will largely take place at CERN. Each of these ‘Centres’ must cater for multiple versions of evolving workflow definitions in an autonomous manner but be centrally coordinated from CERN.

Fourthly, the data collected in the CRISTAL database must be reliably secure (since many processes cannot be undone redone) and available for a variety of purposes. In other words, many different users require access to the CRISTAL data from a variety of viewpoints: construction engineers interpret data using an assembly-oriented view whereas physicists see the detector in terms of a set of electronically-decoded channels and mechanical engineers view the detector in terms of constituent 3-dimensional volumes aligned in space.

Finally, the CRISTAL system must co-exist with other systems during its lifetime. At the outset it must exchange information with a legacy Product Data Manager and, as time elapses, it must act as a source of data for calibration, alignment and maintenance systems, as yet unspecified.

These design constraints cannot currently be satisfied by any commercial offering. The team charged with designing the CRISTAL data model could consequently develop an open and flexible solution. The team adopted an approach in which a model was developed that provided genericity through the use of meta-data, describing general structures, rather than specifying a restricted view onto the workflow or product-specific data. This is in contrast to many workflow solutions which take specific data and attempt to abstract to the more generic.

The particular constraints of time, evolution and viewpoint support quickly led to the adoption of an object-oriented approach to data and process modelling. The UML methodology of Booch, Rumbaugh & Jacobson has been followed; considerable energy has been expended in constructing appropriate Use Cases and exercising these via Sequence Diagrams. The result has been a detailed UML model, presented elsewhere [2].

The main feature of the data model resulting from the use of a generic (rather than specific) approach was its self-describing nature For each class of significance in the data model, a meta-class is defined: e.g part definitions for parts, activity definitions for activities, command definitions for commands and agent definitions for agents. This meta-data mechanism leads to a degree of data independence in that application software is protected from changes in the database schema so that any changes in the specification of, for example, activity definitions does not require recompilation of the accessing software. It is believed that the use of meta-data in workflow and product data specification provides the flexibility needed to cope with their evolution over the ended timescales of CRISTAL production and the flexibility required to cope with ad-hoc workflow specification.

Figure 1 shows a simplified UML object model of the product breakdown structure which underlies CRISTAL. Part Definitions are either Elementary or Composite in nature; composite parts are made up of other parts and the PartCompositionMember object reflects membership of a part in a composite. This meta-data specification is similarly modelled for workflow activity defintions. The next section demonstrates how a meta-data representation can be used to integrate workflow with non-workflow information.

3. The Role ofMeta-Data in Integration

Workflow management systems are increasingly required to co-exist and share information with other non-workflow systems. The issue of workflow interoperability is currently being investigated by researchers and others have adopted a modelling approach [7]. This section shows how a meta-data approach can assist in integrating workflow and non-workflow systems.

Product Data Managers (PDMs) are used in industry to control the documents and data files emerging from the creative stages of product design [8]. PDM Systems provide good support for product documents, particularly at the early stages of design, but their use in supporting the processes in product development is somewhat limited [9]. PDM systems provide no facilities for the enactment or execution of *production* and *assembly* activities. Workflow management systems [10], conversely, allow managers to coordinate and schedule the activities of organisations to optimise the flow of information and operations between the resources of the organisation. Workflow systems are, however, weak at handling the dynamic evolution of process definitions which occurs during the *design* process and can occur even during the enactment of workflow processes. The integration of PDM with workflow management software to facilitate consistency and continuity seems appropriate.

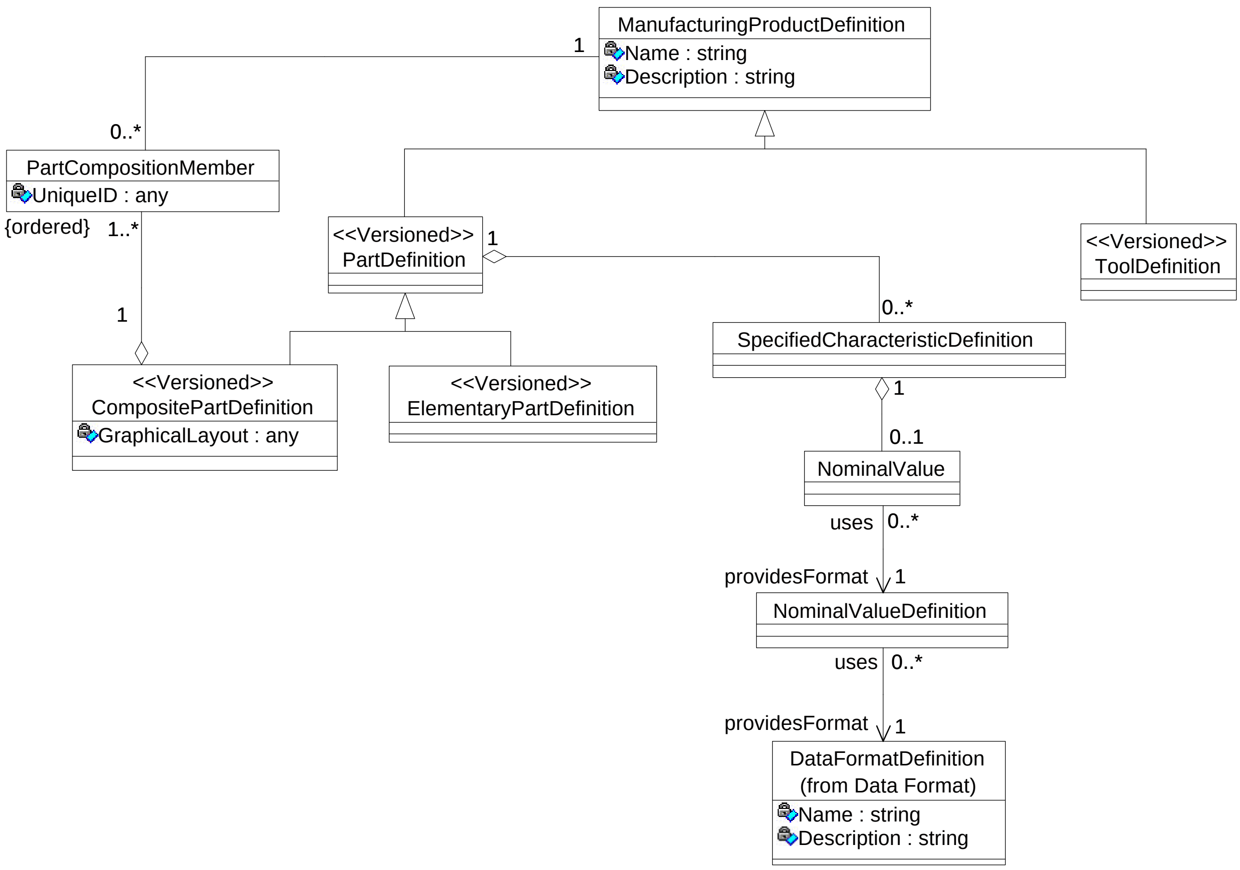

Figure 1: Meta-data for product-related information in the CRISTAL data model

CRISTAL has two distinct functions: one of product data management and one of workflow management. To achieve integration between the PDM and workflow components of CRISTAL, the PDM can be used to store meta-data of parts and activities that need to be executed on the parts. The PDM can then manage the *definitions* of the product and workflow data and the Workflow software can cater for the *instantiation, scheduling and enactment* of those definitions. The PDM acts as the database for the activation and enactment services of the production workflow and handles versioning of the definitions as the schema evolves. CRISTAL therefore controls and tracks parts through the manufacturing life cycle from a PDM-resident design to the Workflow-controlled construction and assembly of the CMS detector.

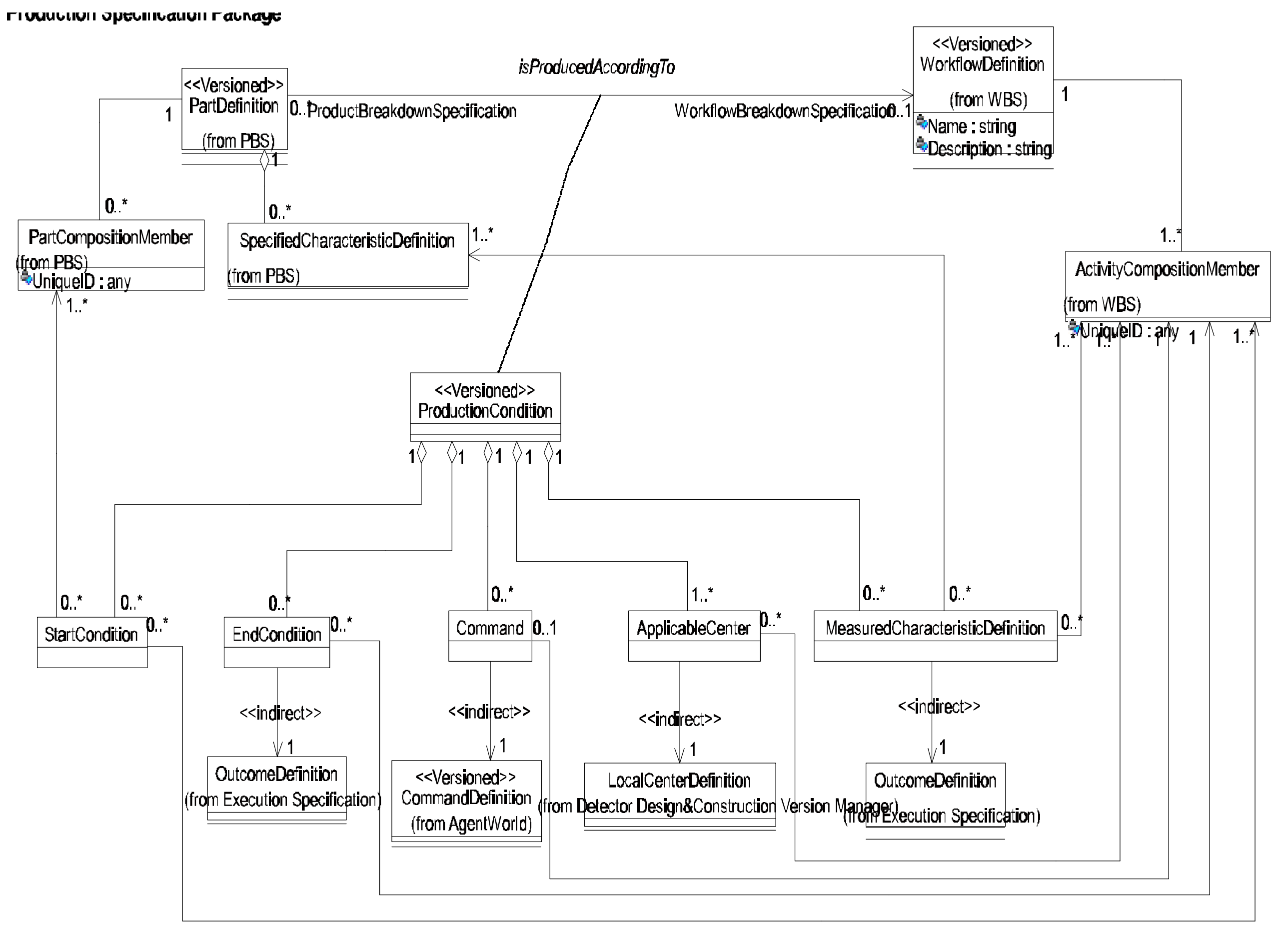

Figure 2: Simplified subset of the CRISTAL model: PDM / Workflow integration and Conditions

Figure 2 shows a simplified subset of the complete CRISTAL data. It shows that there is an association between a given workflow definition and a named part definition (In practice when a part is created in a Centre a workflow object is attached to it). this object being derived from the workflow meta-object.

The CRISTAL data model has been designed so that each assignment of a Workflow Definition to a Part Definition is declared for a specific purpose. Each purpose has associated with it some Production Conditions (see figure 2): the data model captures the definition the *production conditions* required for each assignment of a workflow definition to a part definition. In integrating PDM with WfMS, these are essentially the part objects, commands and locations required as prerequisites for the initiation of a workflow (of a particular definition) on a part (with a given definition). This method of integrating PDM and WfMS through the definition of meta-objects and their mutual assignment is very powerful. It allows other links to be made between aspects of the overall CRISTAL data model. For instance, the same mechanism can be used to assign agents to workflow definitions for the purposes of enactment or the assignment of agents to part definitions for the purposes of data management.

4. Meta-Data and Viewpoints

As the CMS production process evolves so more data, and the relationships between different aspects of the data, must be permanently recorded in the CRISTAL repository. Different groups of users will require flexible ways to find, access and share this production data. The actual information required will depend on the *viewpoint* and the role of the user in the organisation. User groups may well require a maintenance, a survey or an experiment systems management viewpoint. Also, over time, new distributed computing systems will need to interoperate with the older production and manufacturing systems in unforseeable ways. To facilitate the implementation of navigation and data extraction tools, the CRISTAL developers have taken advantage of the object-oriented nature of the CRISTALmeta-model.

Meta-modelling will assist the extraction of data for physics elements provided a *meta-query* mechanism is developed which can navigate the database meta-model, can interpret the structures in the database and can present the data in a form meaningful to the end-user. Meta-queries could be used to assist, for example, in finding data that matched user criteria in analysis programmes. A meta-query facility is currently under development for CRISTAL.

The CRISTAL project has shown the viability and importance of adopting a dynamic object-oriented approach to the development of workflow system software in a rapidly changing application environment. Recently a considerable amount of interest has been generated in meta-object description languages [11]. Work is in progress within the OMG on the Meta Object Facility (MOF) [12] that will manage meta-models which are relevant to the OMG Architecture. Two which are of particular significance to the views in this position paper are the Manufacturing's Product Data Management Enablers [13] and Work Flow Facility[14, 15] meta-models. This meta-modelling approach will facilitate further integration between product data management and workflow management thereby providing consistency between design and production and speeding up the process of implementing design changes in a production system.

The current phase of CRISTAL research aims to adopt this more open architectural approach to produce an adaptable system capable of interoperating with future systems. By modelling the workflow metamodel separately from the workflow enactment runtime model it has allowed the design team to provide consistent solutions to dynamic change and versioning.

6. Acknowledgments

The authors take this opportunity to acknowledge the support of their home institutes. In particular, the support of P Lecoq, J-L Faure and M. Pimia is greatly appreciated.

References

[1] M. Fowler & K. Scott: "UML Distilled - Applying the Standard Object Modelling Language", Addison-WesleyLongman Inc., 1997

[2] N. Baker et al., "An Object Model for Product and Workflow Data Management". Proc of the Workflow Management Workshop, DEXA'98, Vienna, Austria, August 1998

[3] Maes, P. 1987 "Concepts and Experiments in Computational Reflection". ACM OOPSLA 1987 Proceedings. pp 147- 155

[4] J-M Le Goff et al., "Detector Construction Management and Quality Control", CERN CMS Notes 1997/104 & 1998/033

[5] R. McClatchey et al., "A Distributed Workflow & Product Data Management Appn for the Construction of Large Scale Scientific Apparatus". In Proc. of the NATO Advanced Study Institute "Advances in Workflow Management Systems and Interoperability", Istanbul Turkey August 1997.

[6] CMS Technical Proposal. The CMS collaboration, January 1995. Available from ftp://cmsdoc.cern.ch/Tpref/TP.html

[7] C. Bussler., "Metadata in Workflow Management Systems" in Proc of the 2nd IEEE Metadata Conference, Maryland, USA, September 1997 & "Workflow Interoperability - Company-wide Position of the Boeing Company". Draft doc to the OMG Workflow ManagementWorkGroup

[8] M. Philpotts "An Introduction to the Concepts, Benefits and Terminology of Product Data Management", Industrial Management & Data Systems, 1996 4 pp11-17.

[9] P. Pikosz & J Malmqvist, "Possibilities and Limitations when Using EDM Systems to Support the Product Development Process". Proceedings of NordDesign'96 pp165-175,Espoo, Finland 1996.

[10] D. Georgakopoulos, M. Hornick & A. Sheth, "An Overview of Workflow Management: from Process Modelling to Infrastructure for Automation", Journal of Distributed and Parallel Database Systems 3 (2), April 1995, pp119-153.

[11] R. Laddaga & J. Veitch Guest Editors, "Dynamic Object Technology", Communications of the ACM May 1997 Vol.40, No 5, pages 37-69.

[12] Object Management Group Publications, Common Facilities RFP-5 Meta-Object Facility TC Doc cf/96-02-01 R2, Evaluation Report TC Doc cf/97-04-02 & TC Doc ad/97-08-14

[13] Object Management Group Publications, Product Data Management Enablers RFP, Manufacturing Domain Task Force Docs mfg/96-08-01, mfg/98-01-01 & mfg/98-02-02

[14] OMG (1998): Workflow Management Facility - RFP OMG TC Doc bom/97-05-06 & Joint Submission, OMG TC Doc bom/98-06-07

[15] W. Schulze, C. Bussler & K. Meyer-Wegener., "Standardising on Workflow Management - The OMG Workflow Management Facility". ACM SIGGROUP Bulletin Vol 19 (3) April 1998.