Datapipe: A Configurable Oil & Gas Automated Data Processor

Florent Bourgeois[1,2] and Pierre Arlaud[1]

[1]Actimage GmbH, Hafenstraße 3, 77694 Kehl, Germany
[2]Université de Haute Alsace, MIPS EA 2332,
12 rue des Frères Lumière, F-68093 Mulhouse, France
{florent.bourgeois, pierre.arlaud}@actimage.com,
florent.bourgeois@uha.fr

Abstract. Exploration and Production companies need to know where are the oil and gas reservoirs, how much they hold, and whether they can profitably produce oil and gas. Data collection, management and analysis are therefore central to the industry. As in most application areas, raw data are processed, implying several tools and experts interactions. Nevertheless, the oil an gas sector data processes imply unusual scale of, multimodal and long-lived, data alongside with complex analysis. DataPipe is a research project funded by the Eurostars Program of the European Commission which purpose is to develop a platform, toolkit and pipeline for the intelligent, rule-based selection, management, analysis, publishing and display of heterogeneous multimodal data in the oil and gas sector. This paper describes Actinote 4.0, a flexible web-based platform, which is developed to respond to the specified Datapipe context and is dedicated to the creation of specific domain-based process assistant applications that are certified by expert systems.

1 Introduction

Oil and Gas (O&G) are currently worldwide primary resources. Exploration companies perform seismic surveys to locate reservoirs and interpret physical properties of the rock. These surveys generate data which have then to be analysed in order to define the profitability of the reservoir. They can be complemented by aerial or satellite photography, gravitational measures. Plus, test borehole can be drilled to bring data with with positional, radioactivity, temperature, porosity, resistivity and other measures that enhance the geological model. Surveys have been made for years, implying technology evolution. The retrieved data have then changed and their storage also did. Improvements in analytic and drilling techniques as well as shifts in the global economy [1–4] can change decisions, so survey data have a long shelf life. A deposit that was once uneconomic may require a new analysis and interpretation thirty years later. There are problems of maintaining, tracking and accessing very large data volumes for decades, finding and mining old data from different storages, reading and interpreting different media formats and file types. The data are then multimodal, long-lived and difficult to manage. The expected valuable lifetime of seismic data ranges from four to twenty years - up to the point where newer seismic technologies make it cheaper to re-survey

Arlaud P. and Bourgeois F.
Datapipe: A Configurable Oil Gas Automated Data Processor.
DOI: 10.5220/0006163700750096
In *The Success of European Projects using New Information and Communication Technologies (EPS Colmar 2015)*, pages 75-96
ISBN: 978-989-758-176-2

than reprocess the data. As a result, data from many different types, formats and locations have to be found, managed and processed.

Over the past 20 years, geologists have built well databases, geophysicists have developed ways to handle seismic data and reservoir engineers have managed production data. There has been no common approach and the separate initiatives have created solutions with dedicated tools and procedures. Nevertheless, the current trend is to is to move from unstructured collections of physical and digital data toward structured sets of digital exploration, drilling and production data.

Big data and High Performance Analysis are now terms commonly employed to refer to O&G data management in industry white papers [5, 6], and scientific communities[7]. This is due to the massive amount of data it represents, the complexity of the algorithms employed and the variety of data format.

Within this context, Actimage, Dalim Software GmbH, Ovation Data Services Inc. and Root6 Ltd initiated a collaboration around the DataPipe project, which has been selected and funded under the European program EUROSTARS. DataPipe aims for proposing a solution to easily create automated data processor in the domain of O&G which can answer these different prerequisites. The project goal is to develop a platform, toolkit and pipeline for the intelligent, rule-based selection, management, analysis, publishing and display of heterogeneous multimodal data in the oil and gas sector. It will create a flexible system to provide web-based visualisation and decision support based on the analysis of extremely large datasets. The platform will be extensible to big data mining, analysis and display in a wide range of industrial sectors.

1.1 Project Goals

DataPipe will create a new approach to multimodal data management, data mining and presentation, based on process modelling and metadata-based process automation. A new methodology has to be implemented, in order to manage the enormous data volumes, the range of asset types and the processes applied to them in the oil and gas exploration sector. Based on research advances in a number of domains, the DataPipe platform comprises the following elements:

- Intelligent data workflow tools to control the selection and flow of data from multiple sources, their processing and publication on multiple platforms.
- A data selection and management framework, which will deal with the connection to multiple data stores across different APIs and ETL systems.
- A DataAgent processing platform, to connect to archiving systems and tape robots, with a Hierarchical Storage Management (HSM) system to bring files back from storage just in time for processing.
- Access control and security systems to protect the integrity of data from unauthorized personnel or attack to the data store.
- A cross-platform information display system to receive information from the framework and instructions from the job ticket, to tailor the publication.

To implement all of these goals is unsustainable for any of the project partners. Indeed, they require expertise and research specific to too many different domains. This is why each of the firms of the project has to handle a specific part according to their affinities with its themas.

1.2 Collaborating Partners

Being a firm proposing safe and reliable storage systems, *Ovation Data Services Inc.* has specialised in data management for the full life cycle of exploration data in the exploration and production industry. Their services cover all forms of seismic and well log data transcription and migration, plus format conversion, recovery, remediation and de-multiplexing. Their role in the project is to define the domain specific expertise, the existing processes and their results and provide expert feedbacks. Aside of this, they will have to furnish web services to access to their services in order to handle the data all their stages (acquisition, processing, storage, archive, deep storage, destruction) and specify their access requirements.

Root6 Ltd ContentAgent systems address workflow problems in the (digital) film, video and advertising industries. The are designed to support multi-format video media encoding, faster-than-real-time transcoding and streaming transcoding. Their expertise in multiple format video transcoding processes automated and parallelized through the creation of complex configurable workflows is currently limited to local processing. Through this project, they aim to extend the system for Cloud-based operations and adapt their system on the multiple format data found in the domain of O&G to provide faster and automatized data processing based on parallel computing agents.

Dalim Software GmbH specialises in software systems implementing JDF for print, in which XML elements describe all the production processes and material types, regardless of the individual workflow. Their solution enables the creation of processes combining user steps (approval), technical steps and stage steps (milestones) with relevant results stored on databases accessible anywhere through web interface based on the user identified role. They bring to the project their expertise in processes designs implying automated steps and user interactions enabling complex processes requiring human supervision and decisions.

Actimage has been involved in many complex online projects involving cloud-based architecture and has wide experience of combining different mobile technologies. The existing Actinote system is designed to gather and deliver user- and context-centric information for mobile professionals. The system employs scalable Cloud architecture, interacting with multiple data sources (ERP, PIM, etc.), a blackboard multi-agent system to handle data and provide security, delivery and publishing of user related content. Our systems provide a solid expertise for the delivery, security, multi-platform and intelligent interface aspects of DataPipe. Based on this, plus our expertise in mobile and recent web technology applications and requirement based solutions for professionals, our role in the project is to implement user interfaces for the creation of workflows coherent with the specific domain of application that the O&G context represent, mobile devices application for the execution, supervision of decision support based workflows with means of publishing and displaying the information in format tailored to the user needs.

This paper focuses on Actinote 4.0, the solution Actimage developed in the defined context. First it details the goals and requirements specific to the development of a platform for the creation of domain specific based workflow applications. Afterwards it presents the Model Driven Engineering (MDE) approach the platform is founded on. There will then be a presentation of the implementation choices to implement the

method into the Actinote 4.0 platform and a demonstration of the extensions required to implement a reduced O&G data management as a domain specific case to evaluate the solution. Eventually the paper concludes on a review of the results of the project and further work perspectives.

2 The Solution Specification

The previous section highlighted the complexity of O&G data management and the need for a solution to standard their processing which includes complex and time consuming computer assisted activities alongside few user decisions making. This section details the requirement the solution must fill through a stakeholder point of view and then explains how such a specific domain solution can be abstracted to a multi-domain data processing solution addressing a wider range of applications.

2.1 Requirement Analysis

In order to propose such a solution it is important to point out the different actors that are likely to interact together for the generated applications lifetime. We identified four main users in this context.

Software Development Entity. Actimage, as a software development entity aims to propose a solution responding to the domain needs. This implies the generation of application able to handle O&G data by retrieving them from their multiple identified storages and to process them according to data management expert in order to then apply O&G experts analyst specific processes. These processes then result into bankable data that have to be reported and stored for later use.

Considering the number of experts, processes, data storages and data formats, it is not possible to create one application able to respond to any of the processes. On the other hand, it is unthinkable to produce specific development projects for every processes. It would lead to the production of an unlimited number of projects in order to propose to each user the specific process they want to execute with their specific requirement concerning data storages and expected results. It is then obvious that Actimage requires the development of a solution implying an application editor based on software engineering to enable fast and easy development of data processing applications. Thus Actimage will be able to software engineering solutions based on the specific activities of users instead of constraining them to a rigid generic process.

End-users. The end-user brings to the application creation process its expertise on his specific domain of activity, his habits and his expectations on the look and feel of the application. The application process must assist the user in his task and then be a support in his common activities and not a burden. It is then important to give the software architect tools to answer the user needs. It is important to note that the end-users expertise implies specific business vocabulary. hence a gap between the software architect which manipulates software artefacts and the end-user languages naturally

arises. The solution then have to reduce this semantic gap in that it proposes to the software architect artefacts corresponding to the end user's vocabulary. The end-user describes a data management process. This description contains the specific data he wants to use, which implies their format and location, the different algorithms to apply, the decisions to take according to the data and the expected output and output format. A data processing assistant like this becomes pertinent on smartphones if the processing can be outsourced on a distant powerful network of computer and if the application is able to display the pertinent results, enables user decisions and assignation of tasks to other users. Plus if the application present ways to assign tasks to other users, it would be possible to create processes with several end-users interacting to combine different expertises on the data. This requires the solution to handle device to device communication for the assigning and the data sharing activities.

Data Management and O&G Specialists. It is important to note that the generated process if not correctly described can lead to not applicable process in the domain. Indeed, the end-user is an expert in the analysis of specific data. But we mentioned in the previous section that before the data are in the analysis format, they have to be located and retrieved from several data storages, homogenized and made accessible. Due to the disparity of current solutions, O&G experts are used to handle such processes but it is a burdensome task without any of their expertise added value. On top of this, the algorithms of O&G experts analysis processes also imply a set of rules and conditions that only specialists knows. Besides, specification of a process can also entail errors.

Therefore, the different activities composing a full data management process imply activities which present constraints such as specific order (archive data must be the last activity), specific data type as input (only unarchived data are convertible).These specific domains activities being at the center of the processes, it is mandatory that the solution proposes mechanisms to handle them correctly. The specific domain knowledge must be embedded in the application creation process in order to assert that the generated processes are possible and manageable considering the limitation of the system and the sciences of the data management and analysis. Consequently, an analysis of the domain semantic must be performed on the process.

Data management and O&G specialists are then required to define the domain semantic and the set of rules to apply on the system. Their role is then the description of a prescriptive framework to strictly observe. Since several types of analysis can be processed by the different applications, depending on their expertise target, it is of first interest to propose the creation of rule libraries. Each library embeds a specific expertise domain tool, to enable the specific elements of the expertise in the required process creation and avoid to present the unnecessary elements.

Architect. At this point, Actimage solution enables the creation of correct applications in the domain of O&G data management and analysis through the use of an application editor. The person in charge of the application creation is called the architect. The architect has to follow the end-user application description in order to generate an application able to assist him and automize his activities.

The architect goes through two activities. First he engineers his editor environment choosing over the set of specialists libraries and importing the necessary ones. This enables the use of specific domains tools according to the end-user domain and needs. He then performs a process engineering activity creating the application within the editor according to the description given by the end-user. The solution has then to propose tools to compose processes implying end user centric and fully automatic steps. Both must present user interfaces, respectively one to enable user interaction and one for supervision, enabling the end-user to oversee automatic steps evolutions and anticipate upcoming steps.

The different actors interactions described are illustrated in the figure 1.

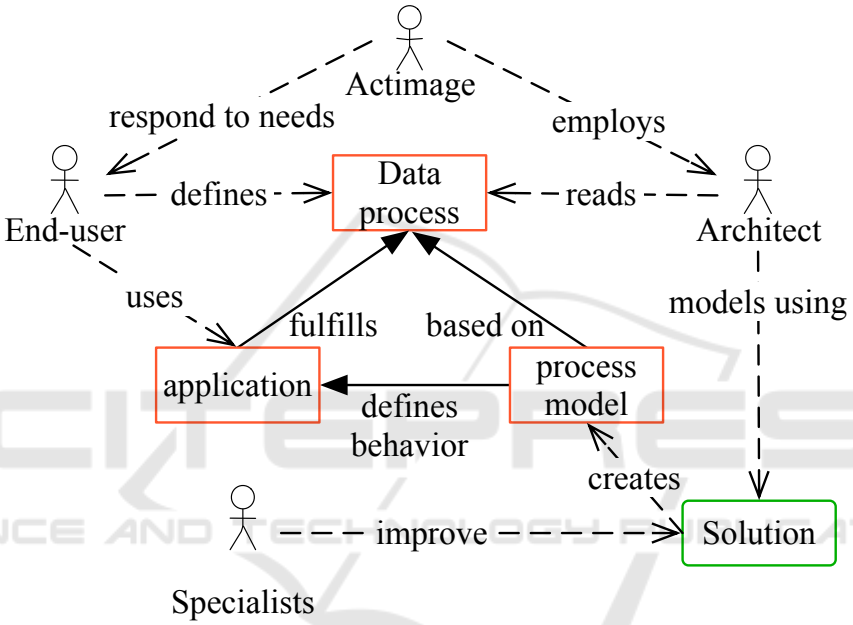

Fig. 1. Solution actors interactions.

To resume, the solution has then to address the following requirements:

- Definition of a complex process composed of automated and user centric activities.
- Execution of process on mobile platforms.
- O&G based processes creation assistant and validation.
- Manipulation of heterogeneous data and data sources.
- Distant processes execution for optimisation and balance of device processing load.
- Role based task assignations to users.
- Web based user interface for the creation, assignation and supervision of processes.

2.2 Solution Generalization

Until now we defined that ACTIMAGE have to furnish tools to create O&G data management processes and to supervise the execution and results of such a process.

It appears from the requirement analysis that the processes are mainly focused on data management and manipulation. Providing that the communicating services are well handled and that the processes are validated by the data management and O&G semantic when created, the format of the data in the application is coherent and the user can manage his work without worrying about the data coherence.

This solution is then based on two specific domains in order to automatize and simplify the expert knowledge based process creation. Since processes are implemented in supervised applications, it is easy to assert that the process status (created, running, terminated) and it asserts that, if the application is correctly described by the end-user, there is no step in the process that can be forgotten.

Such process based applications make sense not only in DataPipe's project context but also in various other domains. In custom-made industry, the prospect phase of meeting the client, understanding its wishes, capturing the environment and estimating the cost of the product would widely benefit of applications that assert that the whole process is performed, that simplify the collection of data either from the firm store to present the products or from the on site visit in order to capture the context of the sale. This domain presents the same requirement as O&G. There are heterogeneous data manipulation, such as camera pictures, measures, notes. Sensors, user interface, cloud-based database are the domain's multiple data sources. That's why all kinds of skills are employed, such as knowledge on the products sold, on the sensors use, or on the price calculation.

It is even imaginable to create applications meant to assist everyday life activities. Indeed, cooking, sport, handiwork are activities that can be represented as processes, requiring user interfaces to guide and assist them, which consume not only data but also materials and create results (meals, health status and manufactured furnitures respectively).

It is then possible to define a common metamodel the three cited examples correspond to which. This metamodel is illustrated on the figure 2.

Thus, instead of a solution based on the verification of two domains, Actimage proposes to define a solution for the creation of multi-domain heterogeneous data handling processes. The processes are validated by their coherence with the domains semantic.

The creation of the solution hence implies several modules:

- domain specific process creation editor.
- heterogeneous data manipulation and presentation interfaces.
- knowledge semantic modeling and verification.
- mobile device applications creation and execution.

The following section presents a MDE approach which was developped to design Actinote 4.0, the generic solution implemented to respond to the mentioned requirements.

3 Model Driven Engineering Approach

As explained in the previous section, the challenges of the Datapipe project can not be resolved by implementing applications for specific user cases unless a whole solution,

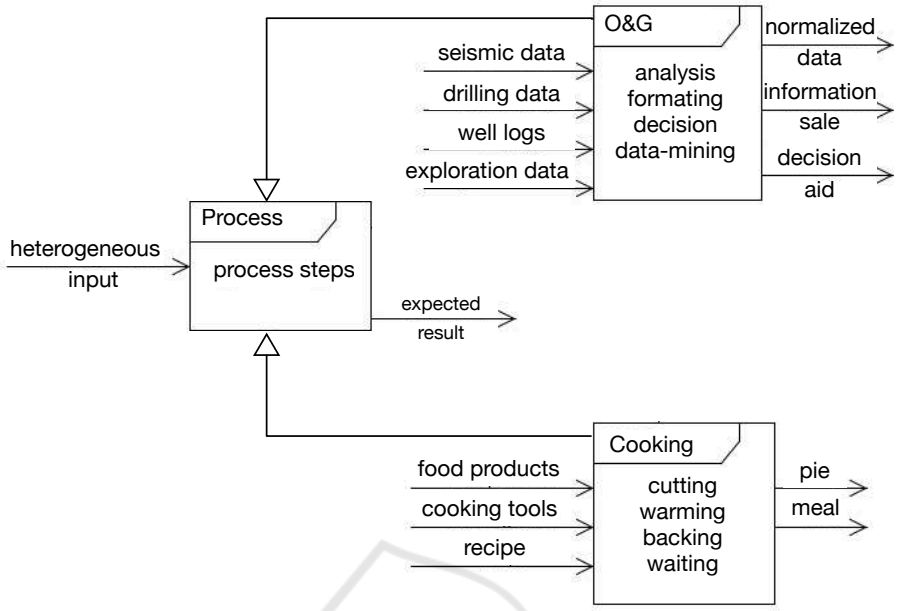

Fig. 2. Processes models and metamodel.

targeting the creation of domain specific data handling processes applications, is developed. This sections describes the structure and producing logic of Actinote 4.0, the solution developed by Actimage.

3.1 Background

Model Driven Engineering (MDE) is a software development method that considers models as the first class artefacts, even considering that everything is a model [8, 9]. Its purpose is to rely on models as development entities and then generate models of lower levels or even code, mapping between models abstractions, model evolutions, system behaviors or applications through the use of model transformations[10, 11].

MDE commonly defines models as a representation of an aspect of the world for a specific purpose. A model never represents the full system, but an abstraction of the system complete enough to represent all the required feature for a given use. A metamodel is a representation of a language able to describe lower abstraction level models. All the models described by the language are conform to the metamodel. This conformance relation thus asserts that the model is constrained by the semantic of its metamodel.

A model transformation takes a model as source and produces another model as target. A transformation metamodel is a mapping between the source model metamodel and the target metamodel.

Surveys [12, 13] proves that MDE, while being a more than ten year old method, is still a recognized method in software industry and several development teams use it in order to approach complex systems development.

OMG's Model Driven Architecture (MDA) [14] is one MDE initiative with a three layer structure. A Computational Independent Model (CIM) describes the business model (e.g. the UML grammar). Then it is transformed into a lower level model through the use of the language it represents. This generates a Platform Independent Model (PIM) which is in our example a specific model described with UML. Last, the PIM model is transformed into a Platform Specific Model (PSM). The generated PSM is the implementation of the system described by the PIM with technology specific to the targeted environment. In our example can be the android application code. Even though our approach does not matches exactly the MDA structure, we will use the CIM, PIM and PSM terms to identify the level of this paper upcoming models.

Using models to specify the system functionalities and then apply model transformations on them, so the implementation is generated, simplifies creation of a group of applications sharing the same description paradigm. It is possible to define a Domain Specific Language (DSL) which is a simple language optimized for a given class of problems[15]. This class of problems is named domain. A DSL enables an easily description of applications in a specific domain using a reduced set of elements. Since the language proposes a reduced set of elements, the model description and mappings are simplified compared to general programming languages, such as C++.

3.2 The Approach Global Structure

As described before, MDE approaches are based on models and their transformations to describe software behavior and automize their implementation based on this description.

Thanks to MDE it is possible to describe a DSL dedicated to the modeling of processes. This DSL is represented by a CIM model. Moreover, the architect editor is based on the DSL. This architect composes the application description with terms extracted from the DSL and thus creates the application process model. This model describes the functionality of the application without considering the implementation specificities. It is then a PIM model. A model transformation consumes the process model afterwards in order to produce the PSM corresponding application.

Nevertheless, the domain specific semantic brought by the specialists' knowledge is complex and implies deep modifications in the DSL with addition of domain specific terms for the architect and, more importantly, of semantical constraints that are hard to represent on models.

Indeed, constraints are often added to the modeling language by the addition of files containing the constraints' descriptions in text such as Object Constraint Language (OCL). This solution presents several downsides. Constraints are placed over objects and object relations, complex constraints are difficult to implement with this approach. This is a problem considering that the domains might be quite complex (e.g. relying on measure semantic). Also, they are in separated files that have to be updated in parallel with the model evolutions. Besides, the semantics have to be analysed on full processes that are only limited by the end-user's description.

In this fahsion, instead of dealing with cumbersome constraints programming, we chose to dedicate the semantic analysis to expert systems detailed later in this paper. We consider that the different domains semantics do not overlap.

To resume, our MDE approach's general structure is composed of three modules and two model transformations that are represented on figure 3. The editor enables the architect to model his processes with domain specific elements, the expert system analyses the domain specific semantics associated to the process to allow only the creation of coherent processes and eventually the application is the final distributed product with which the end-user interacts to execute his process.

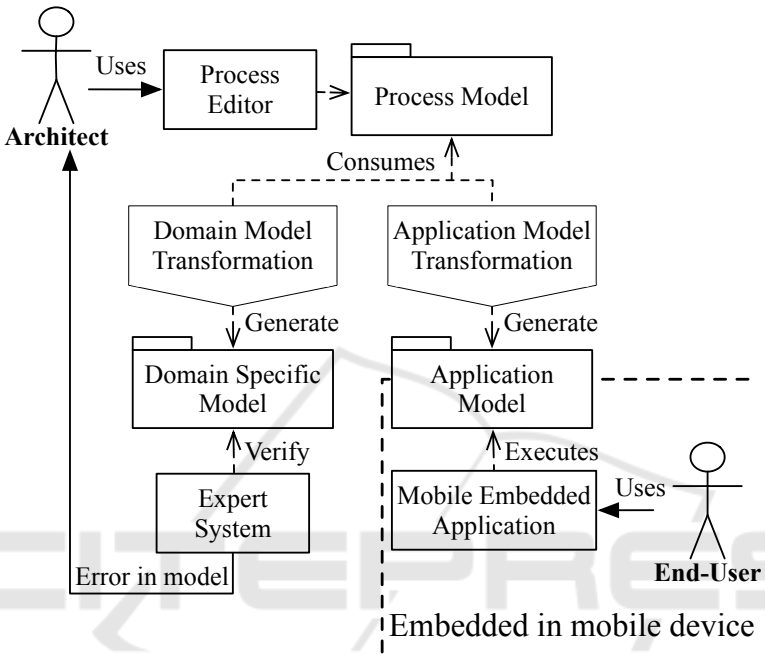

Fig. 3. MDE approach global structure.

This section now provides more details on the different modules composing the approach.

3.3 The Domain Specific Language Definition

In order to be able to create, represent and produce process-based editors and applications, we have to define a DSL able to describe all the possible processes models.

A standard description commonly used to model data processes is the workflow modeling paradigm. Workflows are defined as the automation of a business process presenting several activities, processing any kind of data and connected through transitions[16]. It is a widely used paradigm based on simple elements (activities connected through transitions) defined as being able to represent any kind of process[17]. A large community works on normative use of its elements[18, 19]. In our context, this abstraction can be used as metamodel used to produce the processes representing models.

Looking at figure 2, it is possible to make a direct parallel between the workflow activities, their inputs/outputs and our processes' metamodel. It is also possible to consider a user choice as an activity that transforms two potential futures processes' path

as the one that will be executed. The sole difference is that workflow activities are connected through transitions that can present conditions. These transitions conditions might be associated to either the presence of a correct data (resulting from an upstream activity) or user actions.

Using workflows as a standard representation for our processes' metamodel presents three major advantages. First, it is a simple abstraction that any software architect is used to manipulating, which makes the editor's main elements easy to assimilate. Second, it is possible to propose to the architect complex specific domain elements as simple activities or interfaces. This reduces the semantic gap between the end-user and the architect during the process modeling phase. This enables the creation of processes with less interactions with the end-user to require more precise description. Finally, a lot of workflows editors already exist, for example Datapipe partners already propose solutions based on workflows created through editors. Since those editors create models corresponding to the workflow metamodel, it is possible to use them as an editor approach. For this to be possible, the sole requirement is that the editor can be extended to provide the domain specific elements to the architect and also for the model transformations to be created. Figure 4 illustrates the impact of the use of several workflows editors on the transformation between the process editor and the expert system.

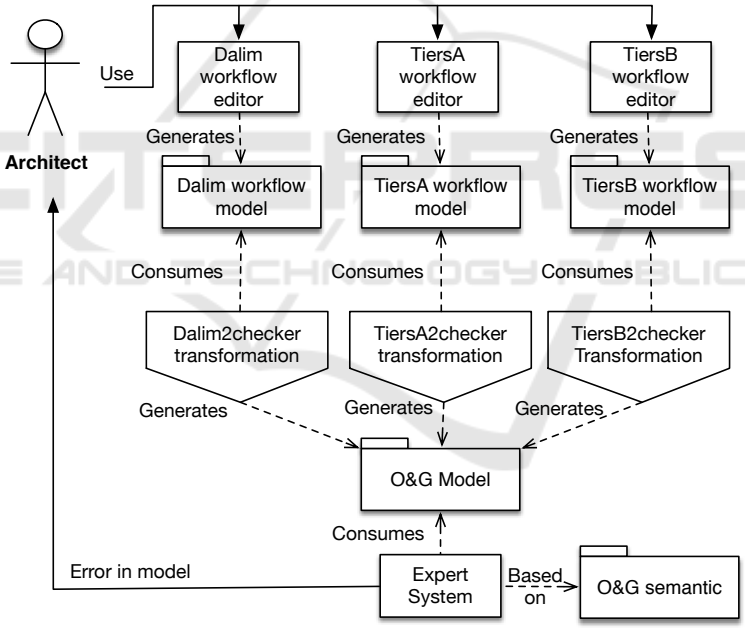

Fig. 4. Use of multiple editors.

3.4 Expert Systems

An expert system, or Knowledge-based System (KBS) is an AI System (input, transformations, output) with several blocks which understands expert knowledge and infers

behaviours to solve a problem in a specific task domain. Expert systems were already used in 1986[20]. They have matured over the years and are now still used especially with the rising domain of ontologies.
There are two types of knowledge[21]:

- Factual Knowledge: Deductions that an expert system should handle as is. Similar to the concept of axioms. This knowledge is widely shared and typically found in textbooks or journals.
- Heuristic Knowledge: The knowledge of good practice, good judgment, and plausible reasoning in the field. It is the knowledge that underlies the art of good guessing.

It is usually said that knowledge-based systems consist of two parts: a knowledge base and an engine. Therefore, as shown on the figure 5, the two basic generic blocks of an expert system respectively have these two responsibilities.

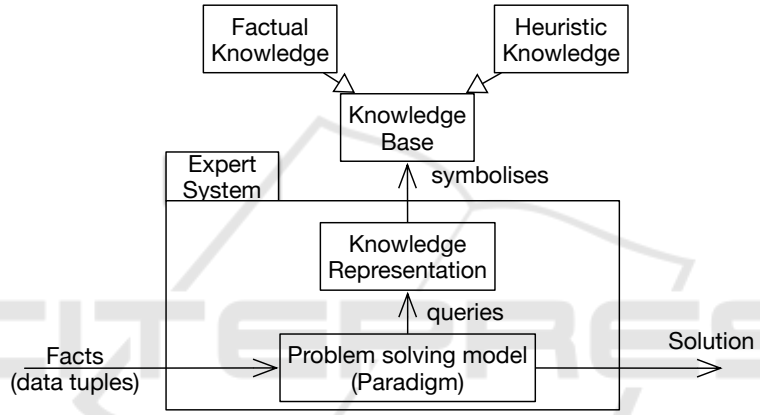

Fig. 5. An expert system.

Thanks to this approach, we are able to ensure that the semantic of the domains will be respected. Such a system, being able to check a domain, is also able to do so with several domains. We did not find examples of overlapping domains, hence our not considering issue of different domains interactions.

3.5 Application

Since our system is meant to generate a family of applications for dedicated domains, it is expected that a end-user requests several applications for different processes. Moreover, the process of posting applications on stores is cumbersome and there is no control over how to access these. The platform must provide a solution to ensure the privacy of the end-user's intellectual property to prevent any unauthorized access to the application, while ensuring the delivery of the application through a simple system.

With this MDE approach, instead of creating a new application for each process model, we propose to translate the model into a PSM which describes the expected behavior of the application and the different interfaces (graphical or to services) that it will use. Then, a unique application will handle any of the descriptions and, based on

an interpreter technology, will execute a behaviour corresponding to the descriptions. The descriptions can be sent to the applicatio through standard push methods on mobile connected devices, automatically providing the new process to the application once it has been created and verified. Combined to a login logic, this allows us to propose a unique application to group and execute any of the processes the end-user requests.

The interpreter executes the application's behaviour according to the workflow activity. Each activity is a milestone in the execution that either starts a process on data or request a choice from the user. The application then only requires to be able to read the workflow and compose interfaces according to the descriptions made in the models. That's why the application must know the editor's different elements in order to be able to interpret them on execution.

Hence, we can propose an application for the different existing mobile platforms which then can handle any of the process models. Which makes the approach able to target multi-platform mobile devices.

Through this section, we presented a Model Driven Engineering approach which enables the creation of a family of multi-domains well-founded processes applications. The next section presents an overview of Actinote 4.0, the Actimage implementation of such an approach.

4 Actinote 4.0 Implementation

The later section presented an MDE approach which answered the complex requirements of Datapipe project. This section presents the Actimage solution Actinote 4.0, implemented following the presented approach. Several details are considered to be out of the scope of this paper due to the industrial nature of the solution. This especially encompasses the different model transformations that will thus not be described.

4.1 The Editor

The editor is the module that is meant to be used by the architect to model the process executed by the application. We stated that such process implies graphical interfaces, workflow activities and data management. The lack of standard processes has encouraged our project team to give users a sense of intuitiveness in the way they can model their activities. The transfer of their operational process into the Actinote 4.0 platform is made accessible with a graphical approach: the architect extracts the flow of the end-user's process in a nodal diagram (which looks similar to a finite state graph) and the description of their constraints.

This normalization and meta-modelling ensures the reliability of the data stores. Not only will the homogeneity of this assemblage facilitate the computational discovery of patterns in the inputs, but it will also allow the utilisation of safeguards based on the specific domains. Since all inputs need to be specified and enumerated, there is in fact no way for the mechanism to be semantically ambiguous. Any incoherence can be spotted beforehand, insuring the integrity of the business knowledge.

The basic metamodel representing the editor DSL is illustrated in the figure 6.

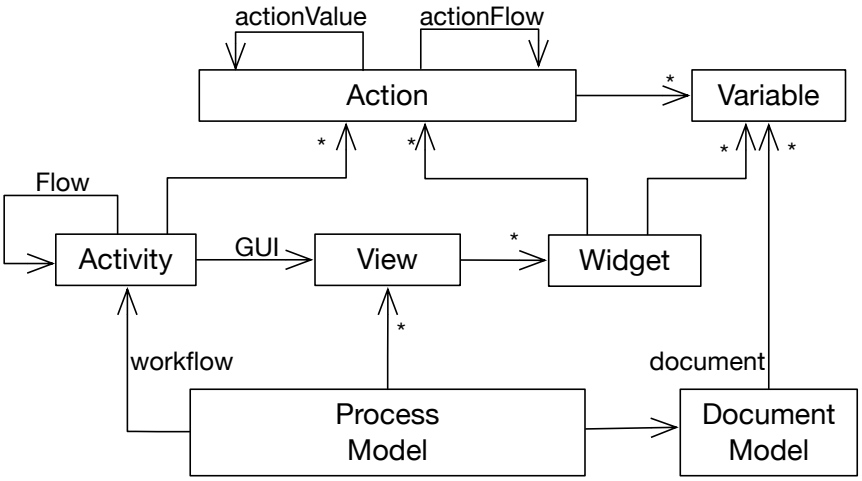

Fig. 6. Actinote 4.0 DSL's metamodel.

With the language represented by this metamodel it is possible to describe any processes. The workflow is modeled as activities chained together through flow links. Each activity presents a graphical user interface composed of several widgets. A widget is a graphical interface element that enables to give access (display and manipulation) to data or make decisions for the end-user. Variables are abstractions of the data manipulated by the process. The document model groups all the data that must be retrieved as the process result in order to automatically generate a report. An action is a specific domain process to apply on a set of data. Some common or complex actions, such as the retrieving of multi-storage data in O&G context, are added to the editor when a specific domain is imported. The architect also can implement specific actions with a nodal diagram dedicated for the data processing. Actions are started either by an activity or by user interactions on widgets.

During our test activities, we observed that the semantic gap reduction brought by the use of our DSL and the abstraction of complex processes as importable actions did not only help the architect understand the end-user specific domain vocabulary, but it also enabled the end-users to edit their own workflows. So our solution, as long as it provides the different complex operations of the process as element of the editor, enables the end-user to model his process model himself.

Therefore Actinote 4.0 is a good fit for the industry because it focuses primarily on the designing of forms and the web-visualisation of analysis results. The whole idea behind this work resides in the opportunity for an expert to be relieved of the time-consuming task load that converting data into a generic form can represent. Thanks to this effort, geologists, geophysicists and engineers can use the DataPipe platform and toolkit to publish and display heterogeneous multimodal data in their realm of expertise. The principle of this responsibility decoupling is that we separate the business logic of the process into three parts: the orchestration of its flow with the activities, the algorithmic aspect of each of its steps with actions and the designing of the display that will provide the users with a mobile access to the process with the widgets. It becomes

thus possible to partition the effort for different employees with different qualifications. Not only will domain-specific experts have the ability to engineer process for virtualizing and structuring production data without requiring any particular skills in software development, but the technical operators and decision makers will be able to run the predefined scenarios independently.

4.2 Workflow Validation

We mentioned in the previous section that the process models have to be validated in order to implement them into applications. This implementation approach adds several users and platforms-based validation requirements. Then, a workflow validation goes through multiples checks:

- Permission and access rights, which may require verifying the coherence of the rights.
- Semantic analysis of fields use and their types.
- Semantic analysis of domain specificities with expert system.
- Vacuity and halting tests (the workflow must have steps reaching a end).
- Responsiveness aberration tests for small displays.
- Consistency checks of the actions graph (which is a set of algorithmic nodes).
- Syntaxic analysis of the actions graph.
- Syntaxic analysis of the activities graph.
- Check of all unused elements (may they be variables types, resources, event graph parts).

Much of these requirements are resolved by the editor's language with typing of actions, variables and widgets and are out of the scope of this paper. We will only detail the expert system validation process.

Rules Engine Implementation of an Expert System. The Actinote 4.0 expert system encodes knowledge in first-order predicate logic and uses the Prolog language to reason about that knowledge. It hence uses a rules engine, which is the most common implementation of an Expert System and based on rules.

The knowledge is represented with a set of production rules Each data is matched to the patterns described by these rules with algorithms such as Rete Algorithm.

The solving entity is thus an inference engine (a.k.a. Production rules system), which uses either forward or backward chaining to infer conclusions on the data.

It's worth noting that, although conclusions are usually implied, their being here inferred shows that we indeed deal with Artificial Intelligence, so the system makes conclusions as humans would.

The figure 7 shows what the expert system thus becomes. Please notice that the Knowledge Base is not explicitly added on the diagram. The confusion between a knowledge base and the way it is represented in our system is usually made on purpose: in a rules engine, we call knowledge base the set of production rules, and not the actual knowledge that the experts have in their brains.

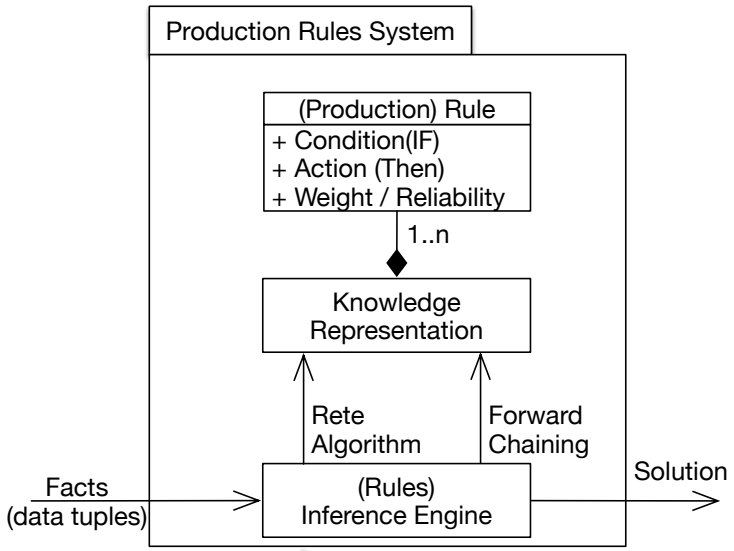

Fig. 7. An expert system based on forward-chaining.

4.3 Data Management

The first target for Actinote 4.0 is to answer to the aforementioned O&G domain data management complexity proposing a front-end to manage the enormous amount of data and their heterogenous nature.

Big-data Choice. Actinote 4.0 has a native support of MongoDB sets of databases. MongoDB is the most used documental database which puts the emphasis on multi-datacenter scalability, resulting in big-data model flexibility and performance. Big-data mining, analysis and display in a wide range of industrial sectors is made possible with this choice.

The data locality of MongoDB instances is an appropriate answer to the needs of O&G data management companies in terms of data. Not only does MongoDB handle billions of documents, but it also sustains hundred of thousands of database atomic operations per second, making it a suitable system for analyzing data. Since it?s also multiplatform, MongoDB can be scattered all across the globe to unite important seismic statistics and pieces of information.

Horizontal Scaling. All the unstructured collections of physical and digital data of the O&G data management companies may be dispatched in structured sets of exploration, drilling and production data. The data can then be split into different shards, meaning there will be different MongoDB servers for different ranges of data. For instance, one may divide the stores geographically and have non-overlapping immutable chunks for each predefined ?location? field corresponding to each area. Considering the built-in geospatial indexes in MongoDB querying system, exploring results of decades of tapes end other capturing data is ensured to remain performant.

Theoretically, there is always the possibility to include a Hadoop framework to solve storage and processing problematics in a distributed way. Computer clusters can thus be accessed to run complex analytics and data processing with Map-Reduce jobs.

4.4 Application

In accordance to the diversity of available media with the Actinote 4.0 software, it's also worth mentioning that it consists on mobile devices of a Qt application, which enables a good homogeneity of resulting behaviors on all platforms. The support of many features such as camera, contact list or network connections are handled in the same way on all platform and the compatibility on most devices (either on iOS and Android but also BlackBerry 10 for instance) remains assured. Another positive consequence of this choice is the integrated ergonomy of the OS: Qt framework adapts to the operating system it is running on so that it can use the standard approach for each graphical component. By doing so, the operators who are running the scenarios can keep the devices they are used to work on and we don?t have to handle resistance to change.

Network of Stores. The structured sets of data are organized in a web of servers and services which are all put together with the cloud computing procured by Actinote 4.0. The uniformity contract of the sets at our disposal can be made practicable by including converters and aggregators of data, or more generally ETL systems, all with the purpose that they are reunified in the beforementioned big-data schemes. In practice, one will firstly design a process, with the benefit webservices and ETL invocation. Secondly, the recorded knowledge will need to be digitized when no virtual save exists in computer understandable formats. Last but not least, this restructured aggregation will be merged and redistributed by means of sharding. This datamining process will maintain the sporadic existence of data with but two main differences: the data will be normalized so by construction rather easy to browse and the interface between all stores will be specified to ensure every piece of information is obtainable on the network. The ontological approach of metrology subjects is a good start for interpreting the production and exploration data which has been performed by Actimage[22].

4.5 Simple Oil & Gaz Implementation

We now present a simple example of the modules modifications implied by the use of a specific domain knowledge. Lets imagine that O&G processes are composed, instead of complex data management activities, of simply four different data manipulation operations: search, format, compare and store.

Search is based on a webservice which returns the data from a selected world area. It can return either numbers or string data depending on the world area (emulate different data storages).

Format allows to format a data into the requested format. If the selected data is already in the correct format, it simply does nothing.

Compress is able to compress a number data.

Store is used to create a uniformed storage. It requires that the data has been compressed.

We will call this domain small O&G in this example.

DSL Additions. In order to handle small O&G processes, the editor requires to propose to the user elements derived from our specification.

The data manipulation operations are actions. A specific action for each of the operations is added in the DSL. Since these actions are obviously distinct we also propose to create one type of activity for each operation. These activities will call the corresponding actions when the user validate their execution.

The Search activity proposes the architect to enter the name of the area in the world to search the data for and returns the result.

Format enables the architect to choose the variable to format, the expected output format (strings or numbers) and the variable in which to store it. It requires the output format and the variable types to be identical.

Compress let the architect choose the variable he wants to apply compression on. It requires the variable to be a number.

Store enables the architect to select the variable to store. It also allows him to choose a storage database to target. This parameter is shared between all stores.

The language also get two type of variables: string and numbers according to the manipulated types by small O&G operations. Compressed data is not a type of data because we consider since the compression is a non destructive operation.

The Expert System. Our expert system is fairly simple because the variable and activities typing validates most of the constraints brought by the domain semantic. Thought, the compressed status of data being not inferred in the process model it is the expert system role to handle it. Also, the type of the data returned by a search action has to he modeled in the knowledge base in order to assert that search actions stores data in corresponding type variable.

The knowledge base is then composed of facts concerning the search areas and rules to verify that search and compress activities are correct according to the expert knowledge. This knowledge being: search activity is always preceded by a compress activity and store activities variable type and storage returned data must correspond. Listing 1 shows the prolog knowledge base. The transformation of process model to specific domain model consumes a small O&G process model and generates a knowledge base extension which contains the different activities, actions and variables facts:

Application Modifications. In order to be able to execute small O&G processes, the application must be upgraded. It is mandatory to provide to the application the code to execute when actions are executed. The GUI widgets composing the small O&G activities are standards validation and data display widgets, hence there is no further development required to adapt the application to the new specific domain.

Listing 1. Small O&G knowledge base.

```
dataArea(string, asia).
dataArea(string,europe).
dataArea(number,America).
dataArea(string,Affrica).

searchVerification(ID) :- searchActivity(ID, Area, VarID),
          variable(VarID,Type),
          dataArea(Type,Area).

storeVerification(StoreID) :- activityFlow(SourceID,StoreID),
          compressActivity(SourceID,_).
storeVerification(StoreID) :- activityFlow(SourceID,StoreID),
          storeVerification(SourceID).
```

Transformations. Both the process model to specific domain model and process model to application model have to be modified. Indeed, the mapping source and target models changed, hence they have to be augmented with the new actions, activities and variables.

Results. It is now possible to use the small O&G editor to create processes, verify them and implement their behavior in a multi platform application. Figure 8 illustrates two processes modeled with an editor. The first one presents an error because there is no compression activity before the store activity. The second one is the corrected version of the first process which is validated by the expert system. Please, notice that figure 8 is an illustration of the process model. It is not produced with the Actinote 4.0 current editor.

This example shows that the addition of a domain specific in the platform induce modification in all the modules of the MDE approach. But, once these modifications have been made once, it is possible to generate as many different processes based on this specific domain asserting that they will be correct by construction. Moreover, as stated before, the editor created becomes easy enough to let the end-user model his processes himself.

5 Conclusion and Further Works

This paper presented that an established fact of the data in the O&G sector is that its interpretation relies heavily on human skill and experience: seismic data can be huge (up to hundreds of petabytes) and full of noise that needs to be manually cleaned. In order to justify the goal of the DataPipe platform, met by cooperating with a variety of specialists in a European project context: to alleviate work that would still be performed by human professionals. After a review of the different project stakeholders requirements, this paper presented Actimage model driven engineering approach to fulfill them. The paper then present an overview of the solution created to apply the approach: Actinote

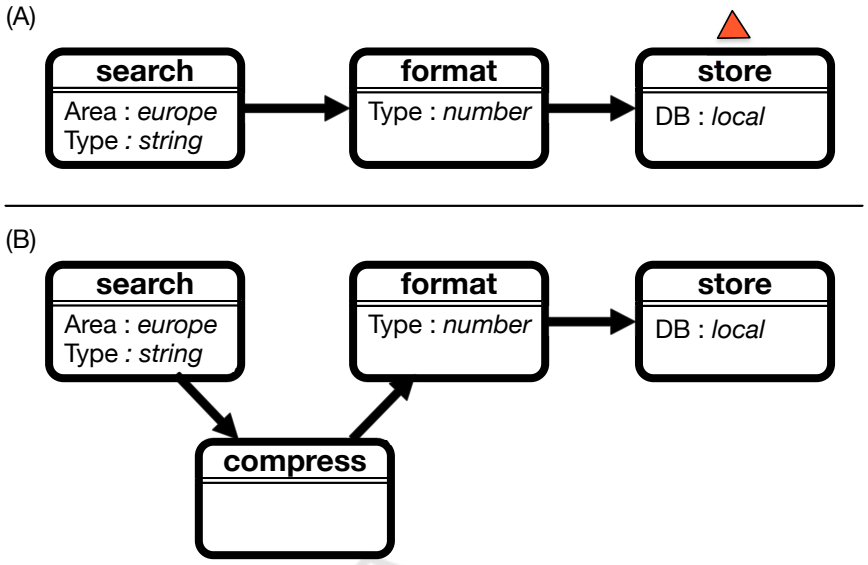

Fig. 8. small O&G processes models. (A) a model with a missing compress activity. (B) a validated model.

4.0 and how it is able to create, validate and implement model specific based data management processes.

Many products have been designed to solve the issues the Oil and gas industry is facing. The adaptivity of the product being a factor of the user acceptance, it seems therefore only clear that filling the gap between the users and the architect, as much as the gap between the noisy content and the normalized format, is essential. Heterogeneousness of formats having been a major subject of the oil and gas field for the past decades, it has been settled in the DataPipe project that giving control to industry specialists was the best approach to counter this environmental disparateness. The interactive Actinote 4.0 platform is the result of cloud-based engineering in that it uses adaptive behaviours to lower expectation differences between individuals and their devices. This brings a flexibility which can be perceived as a catalyst for the support of diverse digital intelligence media. Besides, the seismic stores of contents are arranged in a ubiquitous manner, hence an improvement of adjustability of data for both datamining and analysis purposes. The multiplatform aspects of the UI also mentioned in this article play an important role to the business logic adaptation one can observe using DataPipe software on mobile instruments or displays.

Datapipe project brought to Actimage knowledge in the expert system implementation and use. The size, context and complexity of the project proved to be a perfect opportunity to explore the MDE domain and apply it in, not only generic, but domain specific, user described, processes creation.

The interactions with the project partners allowed Actimage to acquire deep knowledge on specific domains such as the Oil & Gas data management, the print and the film industries. The collaboration in a multicultural context like Datapipes one brought some ideas and solutions that would never have rise otherwise.

Further works implies a deeper relation between the partners to enhance the current O&G knowledge database and then confirm the presented approach scalability to industrialisation. Another current experimentation is the implementation of other specific domain knowledge. Actimage currently works on a metrology based declination of Datapipe. And we also expect to test the two domains combination.

References

1. J. Bielak and D. Steeb, "Abstraction of multiple-format geological and geophysical data for oil and gas exploration and production analysis," Feb. 16 1999. US Patent 5,873,049.
2. R. D. Miller, J. H. Bradford, K. Holliger, and R. B. Latimer, eds., Advances in near-surface seismology and ground-penetrating radar. No. no. 15 in Geophysical developments series, Tulsa, Okla. : Washington, D.C. : Denver, Colo: Society of Exploration Geophysicists ; American Geophysical Union ; Environmental and Engineering Geophysical Society, 2010.
3. Ö. Yilmaz, S. M. Doherty, and Ö. Yilmaz, Seismic data analysis: processing, inversion, and interpretation of seismic data. No. no. 10 in Investigations in geophysics, Tulsa, OK: Society of Exploration Geophysicists, 2nd ed ed., 2001.
4. N. Krichene, "World crude oil and natural gas: a demand and supply model," Energy Economics, vol. 24, no. 6, pp. 557 – 576, 2002.
5. CDW, "High-performance computing in oil and gas," tech. rep., CDW, 2014.
6. HP, "Hp workstations for seismic interpretation," tech. rep., Hewlett-Packard Development Company, 2013.
7. M. Arenaz, J. Dominguez, and A. Crespo, "Democratization of hpc in the oil & gas industry through automatic parallelization with parallware," in 2015 Rice Oil&Gas HPC Workshop, 2015.
8. S. Mellor, A. Clark, and T. Futagami, "Model-driven development - Guest editor's introduction," IEEE Software, vol. 20, pp. 14–18, Sept. 2003.
9. Y. Lamo, X. Wang, F. Mantz, ø. Bech, A. Sandven, and A. Rutle, "DPF Workbench: a multi-level language workbench for MDE," Proceedings of the Estonian Academy of Sciences, vol. 62, no. 1, p. 3, 2013.
10. K. Czarnecki and S. Helsen, "Feature-based survey of model transformation approaches," IBM Systems Journal, vol. 45, no. 3, pp. 621–645, 2006.
11. A. Metzger, "A systematic look at model transformations," in Model-driven Software Development, pp. 19–33, Springer, 2005.
12. J. Hutchinson, J. Whittle, and M. Rouncefield, "Model-driven engineering practices in industry: Social, organizational and managerial factors that lead to success or failure," Science of Computer Programming, vol. 89, pp. 144–161, Sept. 2014.
13. M. Torchiano, F. Tomassetti, F. Ricca, A. Tiso, and G. Reggio, "Relevance, benefits, and problems of software modelling and model driven techniques—A survey in the Italian industry," Journal of Systems and Software, vol. 86, pp. 2110–2126, Aug. 2013.
14. OMG, "MDA Guide 1.0.1," 2003.
15. M. Voelter and S. Benz, eds., DSL engineering: designing, implementing and using domain-specific languages. Lexington, KY: CreateSpace Independent Publishing Platform, 2013.
16. WFMC, "Terminology and glossary," Tech. Rep. WFMC-TC-1011, Issue 3.0, Workflow Management Coalition, Feb. 1999.
17. C. Morley, Processus métiers et systèmes d'information: évaluation, modélisation, mise en oeuvre. Paris: Dunod, 2007.
18. W. Aalst, A. Barros, A. Hofstede, and B. Kiepuszewski, "Advanced workflow patterns," in Cooperative Information Systems (P. Scheuermann and O. Etzion, eds.), vol. 1901 of Lecture Notes in Computer Science, pp. 18–29, Springer Berlin Heidelberg, 2000.

19. W. Aalst, A. Hofstede, B. Kiepuszewski, , and A. Barros, “Workflow patterns,” Distributed and Parallel Databases, vol. 14, no. 1, pp. 5–51, 2003.
20. P. Jackson, Introduction to expert systems. Addison-Wesley Pub. Co.,Reading, MA, Jan. 1986.
21. A. Eardley and L. Uden, eds., Innovative knowledge management: concepts for organizational creativity and collaborative design. Hershey, PA: Information Science Reference, 2011.
22. F. Bourgeois, P. Studer, B. Thirion, and J.-M. Perronne, “A Domain Specific Platform for Engineering Well Founded Measurement Applications,” in Proceedings of the 10th International Conference on Software Engineering and Applications, pp. 309–318, 2015.