A Semantic-based Workflow Choreography for Integrated Sensing and Processing

Yun-Heh Chen-Burger[1], Fang-Pang Lin[2]
[1]AIAI, Informatics, The University of Edinburgh, UK, jessicac@inf.ed.ac.uk
[2]National Centre of High-performance Computing, Taiwan

Abstract— **Real-time automatic and continuous information gathering through EcoGrid offers an unique and immense opportunity for long term ecological monitoring and planning. However, the vast amount of raw data gathered must be dealt with efficiently and effectively so that they may be timely turned into useable information to assist ecological management. In addition, Ecological data tends to subject to environmental changes and exception-prone so that their qualities vary. Adaptive business process modelling (BPM) techniques provide rich conceptualisation to support workflow systems to perform integrated analytical and documentation tasks flexibly and efficiently. We present one such BPM and show how adaptive workflow systems, like those developed at AIAI, can take advantage of Enterprise Models represented in FBPML to provide effective support to users in real Grid environments.**

I. INTRODUCTION

Real-time automatic and continuous information gathering through EcoGrid offers an unique and immense opportunity for long term ecological monitoring and planning. However, the vast amount of raw data gathered must be handled efficiently and effectively so that they may be timely turned into useable information to assist ecological management. In addition, the collected data is often noisy and exception-prone so that analytical methods may no longer operate monotonically, but need to dynamically adapt their behaviours as new results are uncovered.

Enterprise modelling technologies are well recognised for their value in offering structural methods for capturing information in a complex domain. In particular, business process modelling (BPM) techniques provide rich conceptualisation that tends to describe the type of information required by the adaptive workflow systems. However, because of the lack of formal structure in conventional BPMs, their continuous use during system deployment is limited. In addition, their lack of explicitness in describing data semantics prevents them from providing more direct support for workflow execution.

We propose using a formal language within a three-layered framework that allows data to be represented and manipulated explicitly. This language helps to turn the information contained in an informal model into the type of formal model required by an adaptive workflow system. In its current state of development, FBPML (Fundamental Business Process Modelling Language) covers business processes, data models, organisational structure, agents and their capabilities as well as execution logic that gives direct instructions to a workflow engine [6]. FBPML has also been translated to several semantic web languages to provide web services. Currently it has been mapped to BPEL4WS, OWL-S [8], BPML and OWL.

We assist process modellers by giving them a visual modelling language, underpinned by a formal representation and reasoning, that is expressive and easier to use. It lets them specify the information required by a workflow engine. In this paper, we present our formal enterprise modelling language, FBPML. We show how adaptive workflow systems, like those developed at AIAI (e.g. the Task Based Process Manager, AKT Workflow, CombeChem and I-X system [11]), can take advantage of Enterprise Models represented in FBPML to provide effective support to users in real Grid environments.

II. AN ECOLOGICAL CHALLENGE

There was an oil spill at Lungkeng near Kenting, Taiwan, and the Environmental Protection Administration vowed to clear up the pollution and restore the site to its former glory within two months. However, in fact, this is an impossible task - because there had been no ecological data about the area prior the spill and no one knew what the actual "former state" of the site was. In addition, if there had been any research data into the area's ecological system, one would have been able to use it as a basis to seek insurance compensation. This was not an isolated incidence. Traditionally, the aim of nature resource management was focusing on ways to increase commercial utilisation. Examples are arbitrary replacement of mixed forests with trees of higher commercial values, over-fishing and un-regulated use of beaches and oceans, and yet, due to the lack of monitoring and appropriate conservation, it is unclear what effects those actions have on the ecological system as a whole [12].

In 1992, an integrated collaboration project TERN (Taiwan long-term Ecological Research), joining effort with US National Science Foundation's (NSF) long-term ecological research, was formed. Supported by US NSF and Taiwanese National Science Council (NSC), NCHC (National Center for High-performance Computing) have installed and run *Sensor Nets* in several national parks in Taiwan, creating a critical mass of bio- and ecological significant evidences over the past few years.

One example is the 24 hours under water monitoring system at the third nuclear power station near Kenting [1]. Completely undisturbed by natural disasters and human interventions over the last 20 years, the site yields the most prosperous coral reef and marine lives in Taiwan, providing the ultimate standard for

all other marine conservation efforts. Here NCHC have collected information that is of vital importance to both biological and ecological studies that traditional diving method could not have discovered, but it is not an easy task to examine all of the collected data and turn them into useable information.

The first challenge is to process the vast amount of data: one minute of video clip takes 1829 frames and is stored in 3.72 Mbytes. That translates into 223.2 MB per minute, 5356.8 MB per day and 1.86 Terabytes per year for one operational camera, and due to the unpredictability of nature, one may not easily skip frames as they may contain vital information. Based on our own experience, one minute's clip will on average cost manual processing time of 15 minutes: including viewing the clip, labelling the interesting frames, brief annotation and basic classification of the clips. *This means that for one year's recording of a camera will cost human experts 15 years' effort just to perform basic classification tasks.* Currently there are three under water cameras in operation and this will cost a human expert 45 years' time just to do basic processing task. In addition, selected interesting clips may require further analysis; it may also require re-classification if contradictory results have been discovered; comparison studies are also been carried out when appropriate, but those are rather complex tasks and again time consuming. This is clearly a hopeless situation and more appropriate automation methods must be deployed.

It is obvious repetitive tasks such as retrieving video clips, identifying objects of interest in a frame, annotating frames and classifying them in a given structure are monotonic and therefore may be described in a process model and automated through a workflow system. The workflow system may then choose appropriate software according to the task. However, as these videos were shoot in an un-controlled environment, rather than in a lab, they are highly susceptive to weather and environmental changes. As a result, properties and quality of videos may vary greatly even within a very short period that processes that handle them must be extremely sensitive and can adapt themselves according to changes.

It is clear that process modelling techniques and workflow system is helpful, but what are the requirements for such a system? An interesting survey for workflow automation was conducted for the business community by Delphi group in 2002 [7]. Main incentives for a workflow system are: automating repetitive tasks (30%), managing processes and enhancing their performance (25%) and the ability to modify the processes (by the user) without IT personnel's help (20%). Those requirements are relative generic that they share a lot of similarities to requirements within the Ecological domain. We speculate that as their requirements are similar, BPM and therefore FBPML that are useful for the business community may also be useful for Ecologists. Based on more detailed requirements, some of FBPML design rationales are given below:

- Compliant with standardised process modelling languages - so that it is widely accepted and easily converted to other major languages;

- Minimalistic notations - choose a notation that is minimalistic so that it is easy to understand and use while based on well founded theories so that it is extendable when necessary;
- Visual Modelling Language - FBPML provides a visualisation to promote intuitive understanding and manipulation that is absent in other workflow languages;
- Semi-Formal modelling language based on formal theories - that it supports automation and (logical-based) inferencing while provides flexibilities for the user to record informal information;
- Declarative style - rather than procedural style - that it is not limited to a particular implementation style - this also enables a healthy separation between the design and implementation layers;
- Has execution semantics - FBPML has execution semantics, unlike most workflow languages, this enables FBPML to support the creation of a workflow machine at run time;
- Adaptive behaviours at run time - given the execution semantics, dynamic modification of a FBPML process description will effectively change the behaviours of the workflow machine, this ability makes FBPML based workflow system more flexible and can adapt to new requirements;
- Applicable to logical theories - the formal logic based representation captures the necessary types of information to support natural deduction, situation calculus, event calculus, transactional logic, this would enable logical based verification, validation and simulation.

Before we present the language in more detail, the next section describes the three layered modelling framework that over-arching the use of FBPML.

III. THE THREE-LAYERED MODELLING FRAMEWORK

While it is important to use an appropriate workflow language when automating a domain, it is equally important the modelling activity itself is well designed so that the workflow system fit the environment for which it is developed. It is also vital that the process model is not over-prescribed that it dictates and limits implementation options. We have therefore proposed a three layers modelling framework: the objective, logical and application layers. The objective layer describes the objectives, guidelines and constraints for the environment within which the workflow system is to be deployed. Logical layer describes processes to be carried out and data involved - those processes are written in declarative style and may be implemented differently depending on the technologies chosen which are provided in the application layer.

This layered framework has several advantages. It provides an explicit means to describe higher level objectives thus governs the design of a process model. These objectives may also be used as a justification for a design decision that may be reviewed when a process model needs to be modified at a later stage. The separation between logical and application layers allows the modeller to concentrate on higher

level process issues and getting it right, without having to compromise it with lower-level implementation limitations. This separation also allows autonomy for system modules depicted at application layer which may be updated as newer technologies become available.

IV. THE FORMAL LANGUAGE: FBPML

FBPML [6] adapts and merges two recognised process modelling languages: IDEF3 [9] and NIST PSL (the Process Specification Language) [10]. IDEF3 originated from concurrent engineering disciplines and is one of the richest methods available for process modelling. It provides relatively comprehensive visual notations, rich modelling method and model building guidelines. These characteristics make IDEF3 a suitable candidate for capturing processes. Nevertheless, its semantics is informal and its models therefore are open to interpretation.

On the other hand, NIST PSL provides formal semantics for commonly shared process modelling concepts as well as theories that support temporal reasoning on activities. FBPML combines the two different methods by adapting IDEF3's rich visual and modelling methods and mapping those modelling concepts to the formal semantics and theories of PSL, so that reasoning mechanisms and formal analyses can be performed on FBPML models.

In addition, precise process execution logic was added to FBPML so that it is possible to provide direct instructions for execution. As a result, virtual workflow machines may be created and processes enacted at run-time - this was not possible before for IDEF3 or PSL alone. In addition, FBPML makes use of a data language, the FBPML-DL [2], that provides constructs for describing data constructs and becomes an integral part of a FBPML process description. FBPML-DL may also be used on its own to describe a domain(e.g. used as a knowledge acquisition and semantic web publishing tool [3]) and/or used to describe a data model in conjunction with a process model [4] for workflow automation. To provide a brief introduction to FBPML, its process ontology skeleton with a short description is described below:

- Nodes:
  - Primitive Activity: leaf node activities that may no longer be decomposed into sub-activities. Primitive activities are directly connected to application layers.[1]
  - Activity: a general term for all activity types, including composite as well as primitive activity. Composite activities may be decomposed into smaller steps of activities or sub-activities.
  - Role: the role of an actor who carries out the activity.
  - Time Point: a point in time.
  - Junctions: junctions are used to connect multiple activities. They define the temporal constraints and control the initiation and finishing of parallel processes. Each use of a junction is a type of one-to-many (split junction) or many-to-one (join junction) relationships. There are three types

of junctions: AND, OR and XOR junctions.[2]
- Links:
  - Precedence-Link: defines a sequential temporal constraint between two activities.
  - Synchronisation-Bar: defines the synchronisation between two time points.
- Action/Process Decomposition:
  - Task decomposition: decomposes a higher level process into smaller tasks.
  - Alternative decomposition: defines alternative ways of decomposing a higher level process into different sets of sub-tasks; where only one set of those sub-tasks needs to be finished to finish the higher level process.

In addition, there are secondary concepts that are used to specify lower level workflow operations:

- Temporal relations and axioms
- Trigger
- Event (External to an agent)
- Conditional statements
  - Precondition
  - Postcondition
- Life cycle
  - Event life cycle and status
  - Process life cycle and status
  - Data life cycle and status
- Attributes of a process: Instance_Id, Process_type, Life_status, Priority, Average_time_cost, Begin/End_time, Service_Requester/Service_Provider, Trigger, Preconditions, Actions, Postconditions
- Leaf node action types (selected):
  - Data manipulation actions:
    * Create(Instance—Attribute—Relation)
    * Delete(Instance—Attribute—Relation)
    * Update(Instance—Attribute—Relation)
  - Conditional actions
    * Conditional_action
  - User interfaces
    * Report
    * Get_user_input
    * Control of system operation (various functions)
  - Communicating with external agents/brokers (through web services)
    * Construct_issue
    * Post_issue
    * Receive_issue
    * Decode_issue
  - Update object life status
    * Update_object_status
  - Constraint based functions
    * Get_requirements
    * Construct_constraint
    * Check_constraint
    * Get_solution

V. ECOGRID SCENARIOS

Figure 1, 2 and 3 shows screen shots of the *Knowledge Based Support Tool for Enterprise Modelling (KBST-EM)*

---

[1] We do not distinguish between the concepts of process, activity, action and task, these terms are therefore used interchangeably in this document.

[2] The XOR-junction is subsumed by the OR-junction and may optionally be used. We put this notation back to the method to make it more explicit for automation when bridging to (semantic web or web service) methods that do not have OR-junctions.

to describe a simple process model written in FBPML for EcoGrid. Figure 1 depicts a higher level process whose first two activities are composite activities. Figure 2 and 3 detailed the two sets of sub-activities for the two composite activities. In this scenario, the workflow system retrieves video clips made available via EcoGrid. It also chooses the appropriate image processing software for analysing and classification tasks.
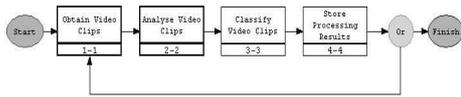

Fig. 1. A higher level process in FBPML in ecology domain

Fig. 2. A FBPML sub-process: Obtain Video Clips
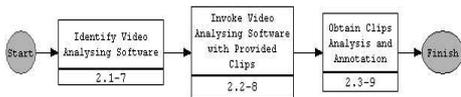
Fig. 3. FBPML Process Model: Analyse Video Clips

As KBST-EM provides logic-based verification and validation, workflow or data related contradictions detected are reported to the user and repairable [5]. A workflow system's dynamic behaviours may be realised and monitored through an *I-X Process Panel*. Figure 4 shows the execution step where video clips are to be analysed.
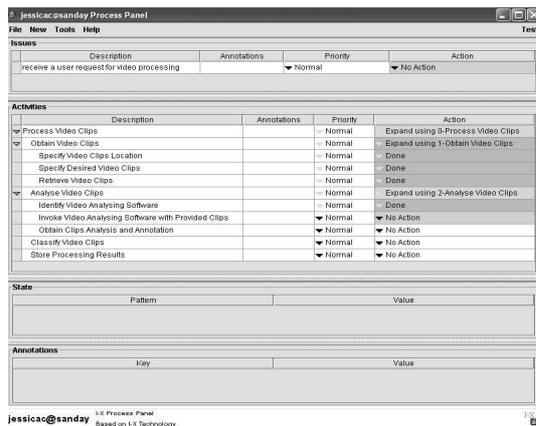

Fig. 4. Executing Analysis of Video Clips in I-X

VI. CONCLUSION

Data collected via remote sensing devices typically is intrinsically dynamic. Their properties and qualities are highly susceptive to external influences such as weather, environmental and biologically changes. Rapid accumulation of vast amount of data via automated monitoring methods makes manual processing impossible. To achieve ultimate use of such information requires sophisticated hybrid approaches: systematic procedures when requirements are monotonic and sensitive and adaptive processing behaviours when exceptions are encountered. After analysing the problem domain in EcoGrid, we proposed a workflow framework that allows us to capture the processes as well as data in a semantically rich environment. This framework is equipped with rich processing semantics that may be interpreted and executed at run time, thus allows us to modify workflow behaviours adaptively based on experiences to suit current circumstances.

REFERENCES

[1] National centre for high performance computing. ecogrid.nchc.org.tw/.
[2] Yun-Heh Chen-Burger. Informal semantics for the fbpml data language. *Informatics Research Report: EDI-INF-RR-0154, School of Informatics, the University of Edinburgh*, October 2002.
[3] Yun-Heh Chen-Burger. *AKT Research Map, AKT IRC project technology show case.* www.aktors.org/technologies/researchmap/, www.aiai.ed.ac.uk/project/akt/, 2003.
[4] Yun-Heh Chen-Burger, Kit-Ying Hui, Alun D. Preece, Peter M. D. Gray, and Austin Tate. Supporting collaboration through semantic-based workflow and constraint solving. *Proceedings of the 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, October 2004.
[5] Yun-Heh Chen-Burger and Dave Robertson. *Automating Business Modelling.* Book Series of Advanced Information and Knowledge Processing, Springer Ver-Lag, December 2004.
[6] Yun-Heh Chen-Burger and Jussi Stader. Formal support for adaptive workflow systems in a distributed environment. *Workflow Handbook 2003*, April 2003.
[7] Delphi Group. *BPM 2002: Market Milestone Report*, February 2002.
[8] Li Guo, Yun-Heh Chen-Burger, and Dave Robertson. Mapping a business process model to a semantic web services model. *2004 IEEE International Conference on Web Services, San Diego, California, USA*, July 2004.
[9] Richard Mayer, Christopher Menzel, Michael Painter, Paula Witte, Thomas Blinn, and Benjamin Perakath. *Information Integration for Concurrent Engineering (IICE) IDEF3 Process Description Capture Method Report.* Knowledge Based Systems Inc. (KBSI), September 1995. http://www.idef.com/overviews/idef3.htm.
[10] Craig Schlenoff, Michael Gruninger, Florence Tissot, John Valois, Joshua Lubell, and Jintae Lee. The process specification language (psl): Overview and version 1.0 specification. *ISTIR 6459, National Institute of Standards and Technology, Gaithersburg, MD (2000)*, 2000. http://www.nist.gov/psl/.
[11] Austin Tate. I-X: Technology for intelligent systems. *www.i-x.info, AIAI, The University of Edinburgh*, 2002.
[12] Chiu Yu-Tzu. Nanwan eco-grid monitors undersea activity. *Taipei Times*, October 2004.