# Workflow Technology for Geo-Processing: The Missing Link

S. Migliorini, M. Gambini, A. Belussi
Department of Computer Science
University of Verona (Italy)
alberto.belussi@univr.it

M. Negri, G. Pelagatti
Department of Electronics and Information
Politecnico of Milan (Italy)
giuseppe.pelagatti@polimi.it

## ABSTRACT

Nowadays GIS users have at their disposal an unprecedented amount of spatial information, thanks to the growing acquisition capacity of the applied survey techniques and instruments, and to the development of Spatial Data Infrastructures and OGC Standards for sharing distributed spatial data. In this context there is the need for new GIS applications that cross the boundary of a single organization and are flexible enough to adapt to the environmental changes. This paper evaluates the applicability of the emerging workflow technology for developing new GIS distributed applications that combine automatic services and human interactions, and are able to deal with large amount of spatial data during long-running processing tasks. Moreover, the limits of this technology when applied to the geographical context are highlighted and some possible solutions to these limitations are proposed.

## Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications—*Spatial databases and GIS*

## General Terms

Management, Design, Human Factor.

## Keywords

Geo-processing workflow, Spatial Data Infrastructure, collaborative processes.

## 1. INTRODUCTION

In recent years the amount of available spatial data has grown considerably: technological advances in acquisition methods and tools have made possible to collect an unprecedented amount of high-resolution and high-quality spatial data, while the proliferation of Internet-based technologies has allowed different organizations to share these data for paying off acquisition and maintenance costs. For instance, modern satellites produce about one terabyte of data per day that are used by several organizations besides the collecting one [5]. A set of organizations interested in maintaining and processing a certain portion of spatial data can constitute a so called virtual organization. The coordination of these new entities is possible thanks to the development of a common Spatial Data Infrastructure, that is becoming a reality in many countries. A *Spatial Data Infrastructure* (SDI) is a technological infrastructure through which different organizations with overlapping goals can share data, resources, tools and competencies in an effective way. Due to its nature, an SDI is usually government-related and regulated by precise rules. In Europe the development of a global SDI is driven by the INSPIRE project[1] that has been recently translated into a European directive. These unavoidable changes pose many additional problems in the storage, access and process of spatial data. The need to share data with other organizations determines new interoperability issues: each organization can store its data in a different format and with a different data model. The increasing size of spatial databases and the processing power needed to manage them exceed the capabilities of a traditional monolithic geographical application that runs on a single server or tightly coupled machines. For these reasons, the Open Geospatial Consortium (OGC)[2] has defined a set of standards for promoting interoperability and sharing of spatial data and services. Moreover, human factors play a central role in the automation of such processes: the knowledge of domain experts is as important as the provided data. Many processing activities cannot be completely automated and the human intervention is required during the computation. In other cases, the development of fully automatic procedures may not be reasonable, even if technological feasible, for instance to manage exceptional situations that rarely occur. In this context, the term *geo-processing* can be used to denote long-running interactive computations that rely on loosely coupled architectures based on self-contained, specialized and interoperable services. The implementation of Geographical Information Systems (GISs) for supporting this kind of geo-processing can be extremely challenging: there is the need to integrate existing systems and to expose them over the network, causing new problems about security and reliability. In addition, these systems have also to be flexible for easily adapting to new environmental conditions and lowering maintenance costs.

Workflow Management Systems (WfMSs) can be good can-

---

[1] `http://inspire.jrc.ec.europa.eu`

[2] `http://www.opengeospatial.org`

didates for tackling these emerging challenges. One of the first attempts to use business workflow technology for geoprocessing is due to Weske et al., in [13] the authors present a system called *GEO-WASA* that allows the integration of data sources and tools for spatial data analysis, querying and browsing through the use of a WfMS. The authors also highlight the benefit of using a WfMS compared with a monolithic GIS system. In [1] Alonso and Hagen present a tool called *Geo-Opera* for supporting the development, execution and management of complex geo-processes. Its major contribution is bringing together, under a single system, functionalities that were previously available only as isolated applications. In the geographical field, the ESRI Job Tracking tool [3] is a proprietary WfMS for modeling and executing geographical processes on top of ArcGIS. It provides the ability to design processes, managing and monitoring their executions, allocating tasks to users and defining user privileges both at task and at spatial data level. While in the open-source context, Foester at al. propose in [4] a flexible client application based on OGC standards for processing distributed spatial data over the web. In [10] the authors evolve their client application by considering the possibility of integrating standardized OGC Web Services using the Business Process Execution Language for web services (WS-BPEL) [8]. The use of WS-BPEL for chaining OGC web services is widely considered in literature and it is suggested by the OGC in its interporability program [9]. A further step towards the integration of workflow technology in the geospatial domain is the use of scientific WfMSs, such as Kepler [7], for chaining geospatial web services, as done in [15], or functionalities provided by existing GIS applications, as in [14]. However, the aim of these works is to obtain automatic processes combing existing tools in an efficient way, no attention is given to the role of domain experts and the collaborative nature of long-running processes.

This paper discusses the maturity of workflow technology (Sec. 2), highlighting the strengths and deficiencies of its application in the geographical field (Sec. 3), and proposes some solutions to overcome such limitations (Sec. 4). The goal is to capture the essential features of an ideal tool for supporting the development and maintenance of new GISs.

## 2. WORKFLOW TECHNOLOGY

In order to face the challenges described in the introduction, new GISs have to meet some new requirements: (1) they have to become *distributed* applications, not only in terms of repositories but also of provided functionalities, (2) they should be *interoperable*, in order to deal with heterogeneous data represented in multiple formats, (3) they have to take a *collaborative* nature, since organizations need to collaborate for satisfying their needs, (4) they have to become *interactive*, because the domain expert knowledge is essential and not all the operations can be automated, and (5) they have to comply with *real-time constraints*, because some information becomes useless if not provided in time.

Workflow technology is increasingly applied for building new information systems, especially when these systems cross the boundary of a single organization. The use of this technology eases the reuse and integration of existing applications, facilitates the deployment of new applications and increases their flexibility, lowering the efforts required to adapt the system to future contingent needs. Given the highlighted requirements, it seems reasonable to exploit workflow technology for building new GIS applications. The term *workflow* is commonly abused and actually denotes different concepts. For our purpose, two types of WfMSs can be distinguished: business and scientific ones.

*Business WfMSs*, recently renewed as *Process-Aware Information Systems* (PAISs), are software systems driven by explicit process specifications, with the aim to coordinate the involved agents in performing their activities. The main goal of a PAIS is to explicit document processes inside an organization, enact them, monitor their execution, and manage human resources. Business workflows are usually classified as control-flow oriented because they focus on describing activities and relationships among them. The usual architecture of a PAIS comprises a Designer, through which a process model can be defined, and an Engine, which generates from the specification a user interface for coordinating and monitoring the process execution. The distinctive characteristic of a PAIS is its ability to interpret high-level process specifications and generate software systems from them, separating the business logics from the software logics: when a change occurs in the organization works, the new processes are documented and translated into operational activities.

*Scientific WfMSs* are software systems developed for automating large-scale scientific experiments. The main goals of scientific workflows are reusing domain specific functions and tools, and supporting their easy integration through a graphical environment, allowing domain experts to define sophisticated analysis ideally without any programming knowledge. Moreover, the nature of the defined computations may require the scheduling on cluster of computers or remote resources; therefore, many scientific WfMSs provide a support for transparently executing their tasks on a Grid environment. Unlike business workflows, scientific workflows are usually classified as data driven: data drive the computation and relationships among tasks are determined by data dependencies; a task can execute only and any time the necessary input is available. This characteristic allows them to support an implicit form of data parallelism: by subdividing the input data into independent chunks, the same task is automatically performed in parallel many times, depending on the available resources.

## 3. THE MISSING LINK

This section analyzes the adoption of business and scientific WfMSs for modeling geo-processes. To ease the discussion it introduces a motivating example and presents its implementation with YAWL System [11] and Kepler [7], respectively.

The process considered in this paper regards the joining of a new spatial data provider to a highly-coupled SDI, where potential participants are known in advance and the global schema for the integrated and shared database has been established with a common agreement [2]. The process includes complex harmonization procedures that require several human interventions on the spatial datasets obtained from the application of geometric algorithms. In the joining process two types of agents can be distinguished: the *SDI manager*, an organization responsible for the maintenance of the global virtual database, and many different administration bodies, called *participants*, that work on overlapping territories and provide their own data for building the overall database. The global virtual database is called Application Database (ADB) and is characterized by a global conceptual schema called Application Schema (AS). The ADB does not physically exist, but it is composed by the union of many Local Application Databases (LADBs), each one maintained by a distinct provider. More specifically, each instance contained in the ADB has an associated owner that is responsible for its maintenance: only the owner can modify or update its instances. Notice that the organization which owns a particular datum can be different from the owner of the physical object (i.e. a building).

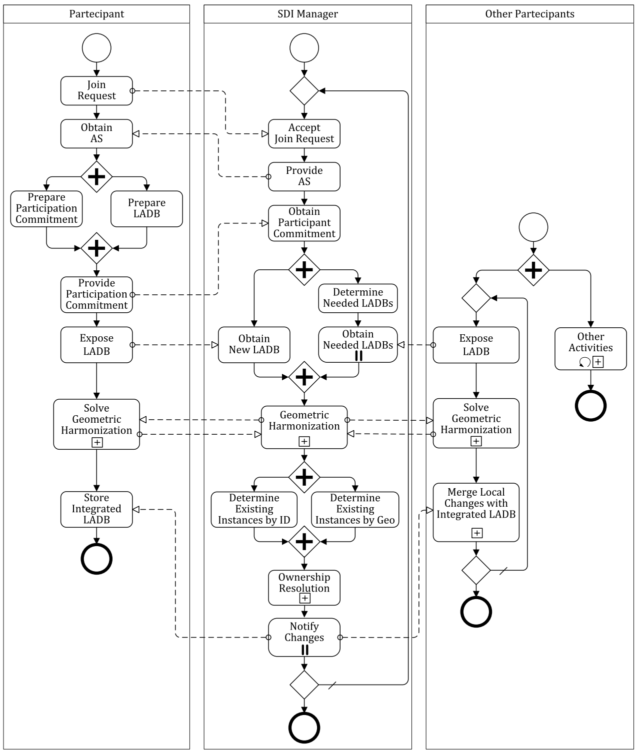

**Figure 1: Process design in BPMN.**

Joining an SDI involves a set of activities which has to be performed multiple times, at least one time for each participant. This process is graphically represented in Fig. 1 using the Business Process Modeling Notation (BPMN) [8]. The process has been decomposed into three sub-processes, placed into different pools, each one representing a role in the joining process. Rounded rectangles denote tasks and the control-flow relations among them are depicted as solid arrows, while dashed arrows represent information flow. Diamond symbols containing a cross enclose parallel activities, while empty diamond symbols conditionally activate one of the underlying branches. A composite task is marked with a plus symbol, while two parallel lines indicate that multiple instances of the task can be executed in parallel and a round arrow that the task will be repeated multiple times.

Referring to Fig. 1, the first activity that an organization has to perform for joining an SDI, is to obtain the AS from the SDI manager. Then the new participant provides his participation commitment and exposes his local data in a way compliant with the AS, as represented by the tasks *Send Participation Commitment* and *Expose LADB* in Fig. 1. The participation commitment contains some information about the provided data, their quality, a set of rules for determining the instance ownership in case of conflicts and the subset of ADB that is necessary to consider during the integration. Once the SDI manager has received the participation commitment, it can start the integration of the LADB provided by the new participant. Firstly, it retrieves the portion of ADB necessary for the integration, as well as the new LADB; then, it starts the geometric harmonization of these datasets. This operation can require several interventions of domain experts for resolving complex situations or correcting the results produced by automatic procedures. In case significant adjustments of the ADB are necessary, if the SDI manager is able to solve itself the harmonization problems, it can provide the automatic procedure with the necessary information to continue the necessary information to continueotherwise the geometry owner will be notified and its consensus for the change is waited. These interactions are represented by the message flow between the *Geometric Harmonization* task of the SDI manager and the *Solve Geometric Harmonization* tasks of both the new and the existing participants. In this activity the competencies of the domain experts are essential for assisting and driving automatic procedures. Geometric harmonization is performed in many refining steps and may be interrupted in several parts waiting for human interventions: it is a long-running process and partial results have to be stored in order to preserve the work already done in case of failures. After the geometric harmonization a check is performed for determining the existence of database instances provided by two different organizations. In particular, for those instances whose existence is independent from their geometry this test is performed on the identifier (ID), while for the other ones the test is geometric and can exploit some information produced during the harmonization phase. If an instance is provided by different participants an ownership conflict is generated and solved by the *Ownership Resolution* task. An automatic procedure determines a possible solution for this conflict, using the information contained in the participation commitments, then the SDI manager can validate or modify the proposed solution. When all conflicts have been solved, *Notify Changes* sends back to all the involved participants the changes, in order to ensure the consistency of the ADB. Besides exposing its LADB, an existing participant continues to perform their own processes, represented by the *Other Activities* task. Since some of these activities can modify the context of a LADB, when an updated version for it is provided by *Notify Chances*, a merge has to be performed between the local changes and those determined by the joining of a new participant. Notice that each successive local update of a LADB will be notified to the SDI manager for continuously ensuring the consistency of the ADB, but this process is not considered here.

The actual process is more complex than the one depicted in Fig. 1: for sake of simplicity many details have been omitted and some aspects have not been included at all, for instance exception handling and compensation. For obtaining a specification sufficiently detailed to be executed in some sort of WfMS, some incremental refinements have to be applied to the process model in Fig. 1. However, obtaining an executable specification from such model is not a simple matter: a huge gap exists between the documentation of operational processes and their executable specifications. The first needed refinement is the definition of data necessary at run-time. The BPMN can be annotated with data aspects, but the data model is primitive and inadequate for describing data-intensive computations, like the geographical ones. Despite this problem, BPMN is commonly used as a graphical representation of WS-BPEL but the mapping between them is far from simple, indeed not all BPMN constructs can be translated into WS-BPEL and vice-versa. For instance, the process model in Fig. 1 will be inevitably translated into a set of uncorrelated WS-BPEL processes, loosing the advantages of an high-level model. The WS-BPEL strengths reside on creating new services by assembling existing ones, but it is not adequate to deal with human interactive tasks, thus ad-hoc implementations are necessary. As a consequence, many WfMSs leverage existing technologies for providing advanced tools and fulfilling the gap. In the following the use of PAISs and scientific WfMSs for implementing the process in Fig. 1 is considered.

YAWL System [11] is an open-source PAIS which is born from an academic research effort pursued for analyzing the existing offerings. This system has been chosen because it provides a clearly stated semantics and captures many of the workflow patterns in a coherent system. YAWL provides a good framework for managing human resources, coordinating their works and monitoring the process evolution, while it is less suitable for performing long-running intensive computations. On the other hand, Kepler [7] is an open-source scientific WfMS developed for supporting time-related intensive computations. It provides multiple heterogeneous computational models that allow the representations of different kinds of systems through polymorphic components, called actors. This system has been chosen because it is one of the most complete and general open-source scientific WfMS and it provides some basic spatial functions. The first part of the joining process, from the initial request submitted by a new participant, to the retrieval of the necessary datasets by the SDI manager, is implemented in YAWL; while the activities related to the geometric harmonization and the ownership resolution is implemented in Kepler.

The YAWL process specification is depicted in Fig. 2. Some decorations have been added to indicate when a task will be performed by the SDI manager (S) or the new participant (P), and the automatic (A) or manual (M) nature of a task. The table legend in Fig. 2 summarizes the basic YAWL symbols used in the example. In particular, an Xor-Split denotes an exclusive choice, only one of its outgoing branches will be enabled on the basis of a particular condition, while an Xor-Join waits the completion of only one of its incoming branches to continue. An And-Split denotes a parallel split, all the outgoing branches will be enabled in parallel, while an And-Join represents a synchronization of all its incoming branches. Finally, a multiple instance task is a task for which several instances of the same type may be generated and executed in parallel. The Designer allows one to specify how each task will be offered or allocated to a certain user and how this task will be started. Some tasks can be marked as automatic, so they will be executed without any human intervention.

Once a process specification is completed, it can be loaded into the Engine. Through a distributed web application, the workflow administrator can manage tasks, assign them to users, monitor their execution, and deal with users and roles. Similarly each user can see its work list, which contains the tasks currently allocated or offered to him and the tasks he is executing, or he can start or complete a task. YAWL can automatically generate, from the specification in Fig. 2, a distributed user interface. However, there is no support for the visualization and management of spatial information. This is a real limitation, since in the geographical field most of the user operations are performed through the interaction with maps. As regards to the roles management, YAWL allows to define privileges only at task level, but in the considered example user privileges have to be defined also at data level: the same task can be performed by different participants but only with reference to a certain portion of the territory. Anyway, the most relevant limitation dwells in its data model: not only spatial types and metadata have to be added, but also the management of data during parallel computations and synchronization phases is an implementer responsibility.

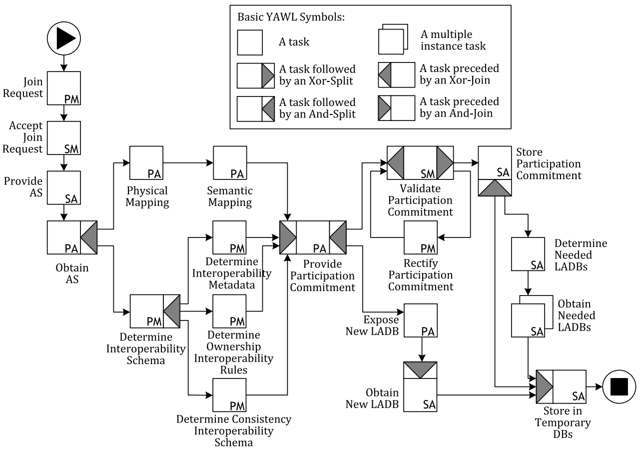

**Figure 2: Process design in YAWL.**

As regards to the computational part of the process, Fig. 3 depicts the process specification in Kepler. In order to exploit the data parallelism offered by scientific WfMSs, the geometric harmonization of the involved datasets is preceded by a *Divide Geometry* task: it subdivides the geometries into homogeneous chunks on which the harmonization can be performed in parallel. The partial results are finally combined by *Combine Results* into a unique output for the subsequent phases. Any time a ownership conflict is identified by one of *Determine Existing Instances by ID* and *Determine Existing Instances by Geo*, a new execution of the *Ownership Resolution* task can be enabled while the search for other conflicts is still in progress. Unfortunately, scientific WfMSs, like Kepler, do not provide any support for managing and coordinating human resources, while the support for human activities is usually very limited. The fundamental assumption in this kind of systems is that an experiment is executed by only one scientist at time and the human interactions are reduced only to provide an input to an automated activity or perform a choice among several alternatives. Therefore, the implementation of the necessary interactions during the geometric harmonization has to be manually solved. Another important limitation it that Kepler is currently provided as a standalone application: for executing a workflow that involves several distributed agents, the user interface has to be decoupled from the engine. For instance, to perform the geometric harmonization, the same workflow instance has to be accessible by all the involved participants, so that each one can solve the problems falling within its competency.

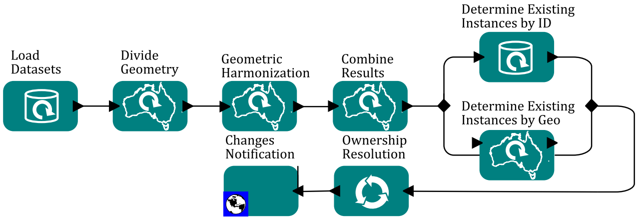

**Figure 3: Process design in Kepler.**

## 4. THE IDEAL SOLUTION

Sec. 3 analyzed the strength and weakness of PAISs and scientific WfMSs as tools for implements flexible GISs. Even if these systems offer invaluable ready-to-use features, none of them seems to be completely satisfactory from an architectural and computational point of view. These limitations regards: (1) lack of a geometric model, (2) lack of user interfaces for the visualization and manipulation of spatial data, and (3) difficulties in exploiting spatial data characteristics during geo-processing. As regards to the first point, spatial data have several unique characteristics compared with data processed in other fields. The main one is the presence of one or more geometric attributes that require the definition of some geometric types and operations. Therefore, existing WfMSs have to be augmented with a geometric data model. The second point is also important, since the visualization of spatial data in forms of maps is an essential characteristic of any GIS application: a numeric representation of spatial objects has little or no meaning for a user, as the information is mainly given by shapes, positions and relations among objects. Users generally perform their operations by manipulating a map representation and not directly the object coordinates. Most of the available WfMSs provide a web-based interface, due to the involved web-based technologies and the deploy easiness. These interfaces have to be enhanced with functionalities for spatial data visualization and interaction through maps: web technology advances make rich web interface possible, but they can expose some limits in managing huge amount of spatial data.

Available WfMSs require many improvements to satisfy the first two points, but the definition of a spatial data model and the manipulation of spatial data on the web are problems deeply studied in literature, the issue is only how to apply and integrate the existing solutions into a WfMS. The real challenge resides on the last point regarding the processing of spatial data. In this kind of processing two forms of parallelism can be distinguished: functional and data decomposition [6]. As highlighted before, PAISs are more suitable for representing functional decomposition of processes, agent coordination and interaction. As regards to the YAWL process in Fig. 2, the first phase of the joining process has been decomposed into several elementary steps among which particular dependency relations have been defined. However, this kind of systems provides little support for data decomposition, which is essential for intensive computations. Optimizations at this level have to be addressed implementing by yourself ad-hoc software components in a general purpose language. Conversely, scientific WfMSs naturally support data parallelism, thanks to the adoption of a data-flow paradigm. They are developed for performing intensive computations and easily exploiting the use of Grid technologies in a transparent way for the users. Moreover, they are studied to simplify the composition of computational blocks and so they offer a better chance to provide ready-to-use functions for geo-processing. However, the adoption of a data-flow paradigm can make difficult to follow the overall state of the workflow execution, because control-flow relations are based on fine-grained data exchanges. Therefore, some problematic aspects remain, for instance how to provide to users a complete overview of the process state of execution, how to seamlessly integrate cross-cutting concerns like resource management, for citing a few.

GIS engineering needs WfMSs able to combine the strengths of PAISs and scientific WfMSs with a model of computation suitable for geo-processing. Given the characteristics of both systems and the peculiarities of geographical processes, it seems reasonable to start from a scientific WfMS and enhance it with the necessary functionalities. For instance, Kepler offers both kinds of parallelism and has been designed for supporting long-running intensive computations on huge amount of data. Available actors are accompanied with an ontology-based description and type interface that together facilitate the creation of new workflows. Finally, it is simpler to enhance a scientific WfMS with the missing functionalities, such as resource management, than enhance the data model underlying PAISs to support data intensive computations. The architecture of a possible WfMS for geo-processing is depicted in Fig. 4. As regards to the extensions needed to obtain a WfMS for geo-processing from a scientific WfMS, the first one is the development of a set of components to handle geographical operations. For instance, Kepler allows one to perform spatial operations by invoking web services and some research has been done for integrating external dedicated applications, like GRASS [14]. However, as discussed in [12], this solution limits workflow transparency, thus it is preferable to enhance Kepler with a core library of GIS functionalities. Many APIs providing geographical functions have been developed, such as the JTS Topology Suite[3]. Anyway, the functionalities offered by these libraries can be currently combined only in a programmatic way; the idea is to use them for developing a set of Kepler actors that can be graphically assembled by users with limited programming skills. Moreover, in order to exploit the data parallelism offered by scientific WfMSs, some specific functions have to be developed for partitioning spatial data into independent chunks of information and combining the partial results. The specific characteristics of spatial data pose some additional constraints in the development of such functions: (1) spatial data do not follow a preconceived pattern, but there can be a considerably

---

[3] `http://www.vividsolution.com/jts/jtshome.htm`

complexity variation across a dataset. (2) Most attention is usually given to the decomposition of data, but in the geographical domain the combination phase is equally important and in some cases even more difficult, since during the reconstruction many properties, like topological relations, have to be preserved. Relatively to the computational aspect, Kepler supports many different computational models, some of them specifically developed for modeling dynamic physical systems. However, in the geographical domain not all these computational models are necessary. We can safely restrict to the one in which components run in parallel exchanging data through channels of ideally unbounded capacity, known as process network [14], and enhance it with coarse-grained control-flow constructs, in order to express the control-flow logic that emerges from fine-grained dataflow relations. This can be useful for mitigating the main drawback of using a data-driven approach with respect to the control-flow one offered by PAISs, namely the loss of information about the overall process execution; e.g. in the model of Fig. 3 one or more instances of *Ownership Resolution* can be executed in parallel with other instances of *Determine Existing Instance by ID*. Finally, the integration of cross-cutting concerns like security and resource management plays a central role. Relatively to this aspect, the engine has to be decoupled from the user interface that will become part of a distributed application for supporting the collaborative execution of workflows by different agents.

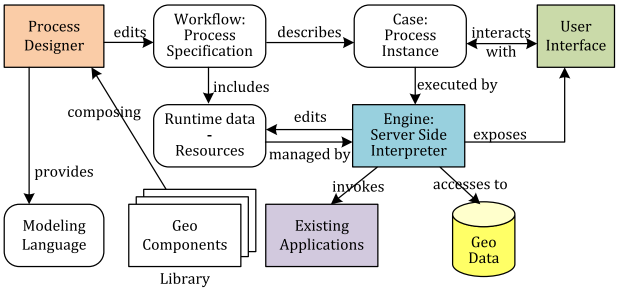

**Figure 4: Architecture of a geo-processing WfMS.**

## 5. CONCLUSION

Environmental changes occurred in the last years have determined the need for a new kind of GIS applications. This paper discusses the applicability of the emerging workflow technology for their construction. It argues that processes in the geographical field can benefit from both the adoption of PAISs and scientific WfMSs, but none of them are completely satisfactory. The found limitations regard three different aspects: modeling, visualization and processing of spatial data. The latter is the most serious one, because it cannot be solved by simply adding features to existing WfMSs. The interactive nature of long-running geo-processing activities, the importance of domain expert knowledge in driving the computation and the need to coordinate the effort of different agents, are better addressed by PAISs. On the contrary, scientific WfMSs can provide a support for intensive long-running computations required by geo-processes. The ideal WfMS for geo-processing has to combine the characteristics of both approaches in a coherent system. In particular, the process network model of computation adopted by scientific WfMSs enhanced with coarse-grained control-flow structures can be the most suitable solution for geo-processing. A planned future work regards the extension of Kepler as highlighted in Sec. 4 for obtaining a WfMS for geo-processing.

## 6. REFERENCES

[1] G. Alonso and C. Hagen. Geo-Opera: Workflow Concepts for Spatial Processes. In $5^{th}$ *Int. Sym. on Advances in Spatial Databases*, pages 238–258, 1997.

[2] A. Belussi, F. Liguori, J. Marca, G. Pelagatti, and M. Negri. Ownership Definition and Instances Integration in Highly Coupled Spatial Data Infrastructures. In $11^{th}$ *AGILE Int. Conference on Geographic Information Science*, pages 379–399, 2008.

[3] ESRI. *Job Tracking for ArcGIS Workflow Management Solution*, December 2007.

[4] T. Foerster and B. Schäffer. A Client for Distributed Geo-processing on the Web. In *Web and Wireless Geographical Inform. Systems*, pages 252–263, 2007.

[5] K. A. Hawick, P. D. Coddington, and H. A. James. Distributed frameworks and parallel algorithms for processing large-scale geographic data. *Parallel Computing*, 29(10):1297–1333, 2003.

[6] R. G. Healey, M. J. Minetar, and S. Dowers, editors. *Parallel Processing Algorithms for GIS*. Taylor & Francis, Inc., Bristol, PA, USA, 1997.

[7] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, and Y. Zhao. Scientific Workflow Management and the Kepler System. *Concurrency and Computation: Practice & Experience*, 18:1039–1065, 2006.

[8] Object Management Group (OMG). *Business Process Modeling Notation (BPMN) 2.0 (Beta 1)*, August 2009. `http://www.omg.org/spec/BPMN/2.0/`.

[9] B. Schäffer. OGC OWS-6 Geoprocessing Workflow Architecture Enginering Report. OGC Public Engineering Report, 2009. OGC 09-053r5.

[10] B. Schäffer and T. Foerster. A Client for Distributed Geo-processing and Workflow Design. *Journal of Location Based Services*, 2(3):194–210, 2008.

[11] A. ter Hofstede, W. van der Aalst, M. Adams, and N. Russell. *Modern Business Process Automation: YAWL and its Support Environment*. Springer, 2009.

[12] C. J. Tuot, M. Sintek, and A. R. Dengel. IVIP - A Scientific Workflow System to Support Experts in Spatial Planning of Crop Production. In $20^{th}$ *International Conference on Scientific and Statistical Database Management*, pages 586–591, 2008.

[13] M. Weske, G. Vossen, C. B. Medeiros, and F. Pires. Workflow Management in Geoprocessing Applications. In $6^{th}$ *ACM International Symposium on Advances in Geographic Information Systems*, pages 88–93, 1998.

[14] J. Zhang, D. Pennington, and W. Michener. Automatic Transformation from Geospatial Conceptual Workflow to Executable Workflow using Grass GIS Command Line Modules in Kepler. In *Int. Conf. on Computational Science*, pages 912–919, 2006.

[15] J. Zhang, D. D. Pennington, and W. K. Michener. Using Web Services and Scientific Workflow for Species Distribution Prediction Modeling. In *Advances in Web-Age Inf. Management*, pages 610–617, 2005.