Report from the NSF Workshop on Workflow and Process Automation in Information Systems*

Amit Sheth, Dimitrios Georgakopoulos, Stef M.M. Joosten,
Marek Rusinkiewicz, Walt Scacchi, Jack Wileden, Alexander Wolf

Abstract

An interdisciplinary research community needs to address challenging issues raised by applying workflow management technology in information systems. This conclusion results from the NSF workshop on Workflow and Process Automation in Information Systems which was held at the State Botanical Garden of Georgia during May 8-10, 1996. The workshop brought together active researchers and practitioners from several communities, with significant representation from database and distributed systems, software process and software engineering, and computer supported cooperative work. The presentations given at the workshop are available in the form of an electronic proceedings of this workshop at http://lsdis.cs.uga.edu/activities/). This report is the joint work of selected representatives from the workshop and it documents the results of significant group discussions and exchange of ideas.

1 Introduction

In the 70's and the 80's, the main objective in using computers in the office was to automate individual work activities. Today, the main emphasis is on capturing and supporting organizational processes that depend on information systems and human resources.

Organizational processes (also called business processes) are collections of activities that support critical organizational and business functions. The activities comprising such a process are tied together by a set of precedence relationships and have a common organizational objective. An important problem in designing and performing organizational processes is effective orchestration of their work activities. This involves distributing, scheduling, controlling, and coordinating work activities among human and information system resources. In addition, effective orchestration must deal with throughput delays, achieve efficient human and system resource allocation, and improve the quality of the resulting products (whether information service or matter).

The need for capturing and supporting organizational processes has led scientists in several computer science disciplines to investigate the issues of understanding, modeling, analyzing, and building processes, and to support them through coordination and collaboration of humans and information systems. In addition, there is an increasing emphasis on distributed technology for reliable and efficient automation of processes that rely on the availability and accessibility of information managed by critical information systems.

The NSF Workshop on Workflow and Process Automation in Information Systems brought together active researchers in the areas outlined above and provided ample opportunities for sharing ideas and exchanging views through workgroup discussions. One of the first contributions of the workshop was the use of the term *Work Activity Coordination* to refer to the emerging field that attempts to combine all these activities. Work Activity Coordination draws from and contributes to multiple disciplines and goes beyond process automation and workflow management. We discuss these in more detail in the following paragraphs.

A workflow process is an automated organizational process, which means that the coordination, control and communication of activities is automated, but the activities themselves can be either automated or performed by people. *Workflow process management* (also termed as workflow management) is the automated coordination, control and communication of

*This workshop was sponsored by the National Science Foundation, under the grant IRI-9528870 (PI: Amit Sheth) awarded by the Database and Expert Systems Program and Information Technology and Organization Program (in the IRIS Division), and the Software Engineering Program (in the CCR Division). All opinions, findings, conclusions and recommendations in any material resulting from this workshop are those of the workshop participants, and do not necessarily reflect the views of the National Science Foundation. Various panel and workgroup discussions were led by members of the program and organization committees- Gustavo Alonso, Omran Bukhres, Ahmed Elmagarmid, Dimitrios Georgakopoulos, Stef Joosten, Marek Rusinkiewicz, Ming Shan, Jack Wileden, Alexander Wolf- and by Clarence Ellis, Gail Kaiser, Adam Porter, and Walt Scacchi. Amit Sheth chaired the program and organization committees.

work, both of people and computers, as its is required to carry out an organizational processes. This is performed by a workflow enactment service element (also called workflow manager, workflow server, workflow engine or simply workflow manager), which is controlled by a computerized representation of the organizational processes, and provides the required services in a computer network. An execution instance of a workflow process is sometimes called a workflow case. Users communicate with workflow enactment service by means of workflow clients, programs that provide an integrated user interface to all processes supported by the system. The technological system by which the organizational process is automated is called the workflow management system.

Researchers in databases and distributed systems have exploited the techniques and tools from database management systems, transaction management, distributed object management and various communication infrastructures, to address the interoperability and coordination issues in supporting organizational processes involving multiple information systems and humans. Workflow management has also been proposed as programming-in-the-large for heterogeneous and distributed information system environments. Researchers in software process modeling and software engineering have developed formal languages for process modeling that have been used to define process analysis, simulation, and execution techniques, tools and integrated environments. They have also developed data collection and data analysis techniques that contribute to continuous process improvement. However, just as the aspects of managing an organization involve many skills and resources, various complementary aspects of supporting and managing organizational processes have been addressed by various disciplines in management, social sciences, and organizational sciences.

Vendors of workflow software have recognized the immediate benefits of organizing both repetitive and non-repetitive activities and developed tools that have been found to be valuable in managing organizational processes of limited complexity in environments that can accept single vendor solutions. Current state-of-the-art technology does not provide a sufficient level of flexibility in supporting processes. For example, changing a process definition locally in small steps is required if different participants are each responsible for their own part of a larger process. Current workflow management systems also do not support simultaneous changes from multiple locations, even when these changes affect a process locally. Other problems include the limited level of interoperability between workflow enactment service elements (or workflow servers) and the information systems they access, lack of transactions, and analysis tools. The capabilities and limitations of the best the current commercial technology has to offer are discussed further in Section 3. New solutions that address these problems will increase the market potential and the real value of the workflow technology. At the same time, it will increasingly affect people's lives, both in the way people perform their work and in the services they consume as a result of other people's work. To accomplish these, further multidisciplinary research is required to understand the interaction between technology, organization, and the human participants, to cope with situations that are complex and dynamic, to learn how to adapt to frequently changing processes, or to heterogeneous environments possibly involving multiple, dynamic, and virtual organizations.

Work Activity Coordination involves such multidisciplinary research and goes beyond the current thinking in contemporary workflow management and Business Process Reengineering (BPR). In particular, instead of perceiving problems in prototypical terms such as the information factory, white-collar work and bureaucracy, we believe that this limited point of view can be explained by a lack of synergy between organizational science, methodologies, and computer science. Multidisciplinary research projects, based on mutual respect and willingness to learn from another discipline, can help to create a thriving research community that builds upon the strengths of different disciplines, such as distributed systems, database management, software process management, software engineering, organizational sciences, and others. This list cannot be exhaustive because people are the main ingredients of any organizational process, and human work raises diverse issues in unpredictable ways. We see the involvement of expertise from multiple disciplines as a necessary component of workflow research projects. We intent to pursue integration of disciplines, relationships between research communities and industry coalitions.

The rest of this report has the following organization. Section 2 provides an overview of some of the challenges facing the emerging interdisciplinary area of work activity coordination. Section 3 reviews the state-of-the-art in commercial technology and discusses related research problems. Section 4 suggests current research areas and future directions. Finally, Section 5 discusses approaches for community building, and possible ways to achieve support for conducting the needed research.

2 Challenges

Research in work activity coordination has resulted to date in both an understanding of the range of organizational processes involved in management and the development of a wide variety of technologies to support those processes. Figure 1, developed during a workgroup discussion at the workshop, gives a qualitative summary of the current state of the art, offering a high-level characterization of how well the technologies are addressing the challenges posed by the processes. The horizontal axis in this figure represents the spectrum of the different kinds of processes found in work activities, while the vertical axis represents a scale of process occurrence frequency or tool type frequency that increases from bottom to top. We can see that most technologies appear at either ends of the spectrum, while most processes appear toward the middle of the spectrum. Therein lies the problem with the state of the art—the existing technologies are not addressing the needs of the majority of processes.

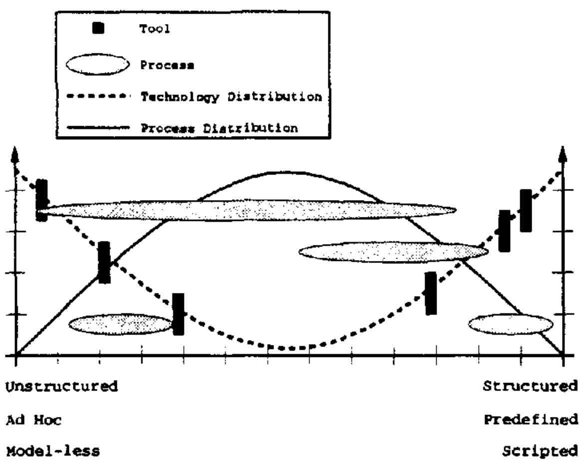

Figure 1: Technological Support for the Process Spectrum

Looking more closely at the figure, we can see that the range of processes can be studied from at least three different perspectives along the spectrum. At one end of the spectrum, processes are unstructured assemblies of activities developed in an ad hoc fashion for which we have no useful models of coordination. An example of such a process is one that might be followed by a hastily formed, cross-functional tiger team charged with solving an emergency problem through brainstorming. On the other end of the spectrum, processes are highly structured activities whose orchestration can be effectively defined before execution and encoded in any one of a number of formalisms, or whose execution can dynamically evolve using well understood and defined rules. An example of a highly structured process would be a formal quality review conducted at regular intervals on a mature product. Another example is that of a supply chain model (used in a recent WfMC interface level 4 interoperability demonstration) that involves scenarios such as 1) retailer replenishment direct from the distributor, 2) retailer replenishment via third party warehouse, and 3) retailer and distributor stock replenishment, including the need for customs documentation from the trucking company's bonding department.

In practice, many processes share characteristics of both ends of the spectrum. Moreover, the individual processes tend to contain elements that widely span the spectrum. Therefore, it is rare to find purely unstructured processes, just as it is rare to find ones that can be characterized as purely structured. This is represented in Figure 1 by the horizontally elongated ovals.

Current technologies do not match this distribution of processes. Most technologies address the processes found at either end of the spectrum. For instance, group decision support systems that can be found in meeting rooms are aimed primarily at supporting unstructured, ad hoc collaborations. As such, they do not attempt to embody a model of coordination, since any such model would in general be inappropriate. Workflow management systems typically involve the use of a scripting language, rules, or dependency specifications, and thus tend to serve highly structured processes best. Overall, in terms of Figure 1, current technologies can be said to be point or 'stovepipe' solutions and are consequently not effective at supporting wide-spectrum processes.

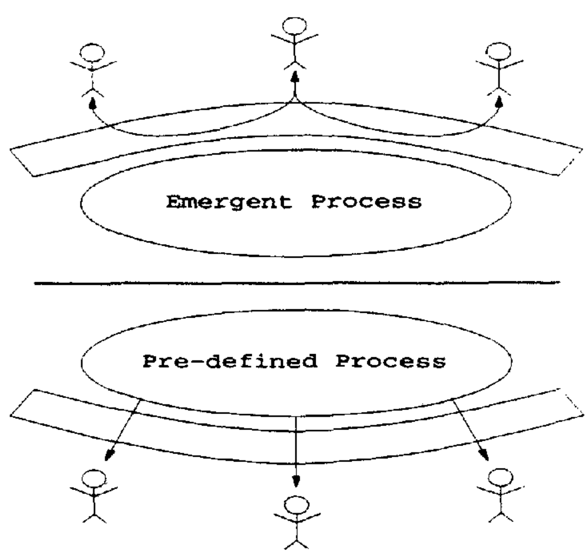

Figure 2: Two Human Roles in Process Management

An issue orthogonal to the above is the role of the human in defining, executing, and evolving the processes. Figure 2 offers two opposing views of this role. In the first view (top of Figure 2), the process as a manipulable entity emerges from the interactions among humans. The technology serves only to mediate, and perhaps also record, the interactions, but does not otherwise coordinate or control the activities of humans as they perform a process. In the second view (bottom of Figure 2), the human interaction is explicitly coordinated and controlled by a predefined process. Changes in the interaction must be made through an evolution of the defined process. Thus, one perspective on the difference between these two views is that it represents a question of who is in control — the process or the performer. Another perspective is that it represents a question of where the evolution in a process takes place, implicitly in the actions of the performers or explicitly in the process definition. Real-world processes involve both views.

Thus, the challenge for the emerging work activities coordination community is to discover ways to incorporate human work in the context of both wide-spectrum processes and technological infrastructure, for the purpose of realizing highly productive organizations in which there is integrated coordination and control of both human and automated activities. This raises a mix of technological, organizational, and methodological issues that must be addressed by radically new research approaches and initiatives.

Technological challenges vary widely, involving issues such as dynamically changing processes, heterogeneous computing environments, exceptions, interoperation of processes, reuse of process definitions, static and dynamic analysis techniques, and more. Which technology is needed in different types of processes is not well understood. For example, a major issue is whether basic or extended transaction models can serve as building blocks in a system that coordinates work activities. Interoperability issues arise because workflow management systems are used within existing heterogeneous environments; protocol definitions, programming interfaces, data sharing primitives, and human interface environments all need to be addressed as interoperability issues that are relevant for integrating work into organizational processes. Research in data communication, especially at the level of CORBA (Common Object Request Broker Architecture) objects and the OSI application layer, is necessary to incorporate available multimedia, groupware, EDI (Electronic Data Interchange), messaging, and Web technology seamlessly into work processes. These are discussed further in Section 3.

Organizational research is needed to understand the problems and expectations of the user community in workflow management. This is currently a vendor-driven discipline that has set high expectations inside and outside the user community. Workflow technology is expected to do for information based ('white collar') work what automation has done for material based ('blue collar') work. The user and vendor communities signal the need for carefully documented case studies and evidence about the impact of activity coordination on the organization. Organizational issues are diverse in nature, putting forward problems such as inter-organizational collaboration, cognitive distance between humans and process representations, performance measures, and practical limitations on workflow management systems. Today, we still lack a detailed understanding of how to support the richness of real organizational processes, which must be understood in many different dimensions: structured versus unstructured; synchronous versus asynchronous; single instance versus multiple instance versus massive instance; and human versus machine.

Methodological issues arise when bridging the gap between technology and organizations. We currently suffer from confusion caused by weakly defined concepts and a lack of consensus about the way in which these concepts are used. The situation is comparable to the situation of database technology in the 1960s, which was resolved to a large extent during much of the 1970s and 1980s by the definition of the relational model. The glossary of the WfMC has not resolved the problem because it contains narrative definitions only, which are incomplete and at times ambiguous. To meet this challenge, a careful analysis of concepts is needed. This can lead to a well-defined framework of concepts with operations that integrate the concepts in which organizations express their problems with respect to work processes. Methodological issues thus include the representation of process definitions, modeling of coordination and control, understanding which methods and techniques apply in which situations, and the consistent use of concepts. This understanding enables the development of effective techniques to analyze and design systems to support various types of work processes. The technology needed to support this is method engineering technology, so as to experiment with various modeling techniques. Also, process definition languages, static and dynamic analysis techniques, and organizational metrics are required components of the solution to this issue.

3 Evaluation of the Current State of Technology

Seen from the distributed systems and database perspective, the current state-of-the-art in workflow management can be characterized by considering the degree to which a workflow process depends on humans or software for performing and coordinating activities. Such a characterization is depicted in Figure 3.

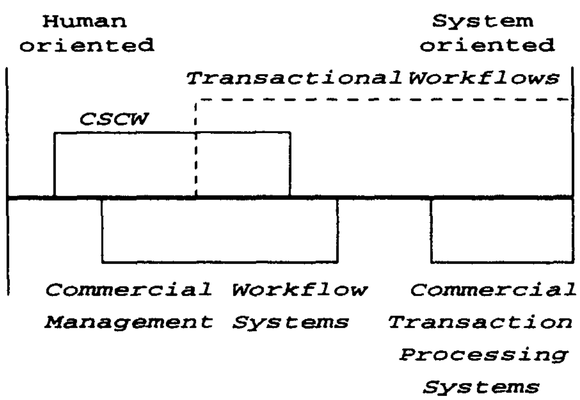

Figure 3: Technologies with respect to Software and Human Supported Activities

On one extreme, a workflow process involves human collaboration while performing and coordinating activities, with relatively few or no automated activities. To support this type of process, a workflow management system must provide distributed systems and database infrastructure for the coordination and collaboration of people.

On the other extreme, a workflow process involves coordination of heterogeneous information systems and workflow applications that perform highly automated activities with relatively little or no human participation. Consequently, a workflow management system must deal with data and information system heterogeneity, distribution, and autonomy. In addition, such systems must match workflow process definitions to the functionality and data provided by legacy heterogeneous information systems and/or their applications. They must provide adequate throughput, and ensure consistency and reliability by providing services for concurrency control and recovery at the level of activities, but also at the process level. Issues such as workflow definition, workflow specification language, monitoring and analysis tools, exception handling, client and server mobility, prioritization, and deadlines must be addressed in all types of workflow management systems.

Also depicted in Figure 3 are the relative positions of related technologies. CSCW overlaps with workflow management in situations where workflow processes involve predominantly human activities. Commercial transaction processing systems overlap with Workflow Management Systems when workflow applications are submitted as database management system- or transaction processing monitor-supported transactions.

In the following sections we discuss the features and capabilities currently supported by workflow management systems. These include commercially available products and home-built applications currently in use in various organizations. In addition to the current-state-of-the-art, we discuss sample problems for further research.

3.1 Process model

Workflow management systems typically support activity-based models, although a small part supports a conversation-based model. Conversation-based methodologies stem from the Winograd/Flores Conversation for Action Model, which is based on speech act theory. A conversation-based methodology reduces an organizational process to a network of commitment loops consisting of four phases based on conversation between a customer and a performer. The resulting organizational process reveals a social network in which a group of people, assuming various roles, fulfill an organizational process. An activity-based methodology reduces an organizational process to a network of activities and subprocesses. The resulting graph represents a control flow network of organizational activities.

Activity-based methodologies focus on modeling activities instead of modeling the commitments among people. Activity-based process models typically support the following concepts:

- *process*: a set of activities (also termed as tasks or steps) and precedence relations.
- *activity*: a unit of work that an individual, a machine, or a group can perform in an uninterrupted span of time. The execution of one activity consists of a sequence of interactions (called events) between the performer and the workflow management system, and a sequence of actions that change the state of a particular instance (or case).
- *information object*: e.g., documents, data records, images, phone calls, fax, printouts, etc.
- *role*: a placeholder for a person, group or an information system service related to a particular activity.
- *agent*: person, group or an information system that fills roles and interacts while performing activities in a particular workflow instance.

To provide different levels of abstraction, workflow management systems typically support the nesting of processes. Higher levels of abstraction help in capturing the organizational process as it relates to the operations and the organization (sub)units that participate in carrying out the process. Modeling at these higher levels is typically devoid of implementation, technology or software details, organizational process. This and is sometimes called organizational/business process modeling. The lower levels of abstraction are required to capture a variety of details related to support of the process by a workflow management system. Modeling at these lower levels includes details about the involved information systems and other software, and the allowable event sequences. This implementation and technology oriented modeling is sometimes referred to as workflow modeling. The issue of level of modeling and use of abstractions is not yet well understood, and when it is not necessary to distinguish between the levels of modeling, we use the term process modeling.

Activity-based process models do not always support the explicit specification of what it means for a workflow process to accomplish its objectives at all levels of abstraction. For example, this includes determining which activities must be completed for the workflow process to be considered successful, and dealing with the exceptions generated as a result of accessing various information systems in the process. Research in the area of distributed transaction management and extended transaction models addresses the issues of correctness and reliability at varying levels. However, whether to or how to adapt process models to accommodate non-transactional activities and more heterogeneous agents such as humans and legacy systems, and not just database management systems and transaction processing monitors, is an outstanding research issue.

3.2 Process/workflow specification language

State-of-the art process/workflow specification languages support the specification of the following:

- activity structure (control flow) and information exchange between agents (dataflow) in a workflow process, specifying that activities can be executed in parallel or that an activity needs to wait for data from other activities
- exception handling, specifying what actions are necessary if an activity fails or a workflow process cannot be completed
- duration, specifying initiation and completion time of an activity
- priority attributes, specifying priorities for activity scheduling

Most workflow management systems provide graphical specification of workflow processes. The available design tools typically support the iconic representation of activities. Definition of control flow and dataflow between activities is accomplished by connecting the activity icons with specialized arrows specifying the activity precedence order and their data dependencies. In addition to graphical specification languages, many workflow management systems provide rule-based or constrained specification languages.

State-of-the-art languages are unstructured (e.g., they are based on graphical 'node and arc' programming models) and/or rule-based. Unstructured specification languages make debugging/testing of complex workflow processes difficult. Basic notions of structured programming languages and corresponding GUIs have to be developed to improve workflow specification testing and debugging. Rule-based languages become inefficient whenever they are used for the specification of large and complex workflow processes. This is due to the large number of rules and overhead associated with rule invocation and management. Development of more efficient specification languages, with attention to testing for correctness and completeness of specification, is required to support large and complex workflow processes.

3.3 Testing, analysis, simulation and animation tools

Workflow testing tools simulate a workflow process by allowing input of sample data and external events such as activity completion, deadline expiration, and exceptions. Analysis is needed to uncover logic errors and to obtain (static) estimates of completion times. Animation is done to demonstrate a workflow process on a computer screen, which is evaluated by stakeholders during design. Simulation gives insight in dynamic behavior, stable vs. startup behavior, location of possible bottlenecks, etc. Workflow analysis is done by taking into account workflow execution or simulation statistics. For example, analysis tools can gather statistics on performance and suggest alterations to the workflow specification to improve efficiency. Some contemporary products supply simple testing tools, but most are inadequate.

The sophistication of testing and/or analysis tools has a direct impact on the development speed, and the ease of specification and implementation. Additional research is needed to introduce such workflow-specific tools. Support for simulation and animation is desirable, and there is little work that supports any comprehensive model for realistic environments.

3.4 Workflow monitoring and tracking tools

Workflow monitoring tools can present different views of workflow process execution. They illustrate which activity or activities are currently active, by whom they are performed, the priorities, deadlines, durations, and dependencies. Administrators can use such monitoring tools to access statistics such as activity completion times, workloads, and user performance, as well as to generate reports and provide periodic summary of workflow process executions. Other applications include discovery of process models from process data, and validation of process models against actual process executions.

Tracking can provide valuable statistics and traces for completed activities. Open problems include the definition of useful process metrics and the integration of automated and manual data collection techniques.

3.5 Systems architecture, scalability, performance, and component redundancy

Some state-of-the-art workflow management systems have open client-server architectures and complete application programming interfaces (APIs). This allows everything that can be done through the user interface also to be done via an API. In addition, the API can be used to introduce specialized user interfaces designed to meet specific application requirements. Loose coupling is achieved by communicating work items. A work item is a message that represents the occurrence of an event in the entire system. Receiving a work item provokes a reaction as though the receiver notices the event. The rules by which work items are routed, kept, distributed, and processed are not embedded in the individual activities, but are kept together in a repository under control of the workflow management system. The separation of control structure from process structure is a characteristic of workflow management systems.

State-of-the-art workflow management systems typically cannot support more than several hundred workflow instances per day, and the use of more powerful computers often does not yield corresponding improvements in throughput. Some processes require handling a larger number of workflow instances which may exceed the number of transactions that transaction processing systems can handle. For example, telecommunications companies currently need to process tens of thousands of service requests a day, and even several thousands per hour during peak hours. A paper at the workshop discussed high throughput requirements for a workflow process in a genome lab. To date, workflow management systems cannot handle such workloads. These problems are due to immature workflow management system server technology, limited server scalability, and lack of component redundancy for dealing with server failures and/or load balancing. Although there are workflow vendors that are currently attempting to address some of these problems, further research is needed to extend the current state-of-the-art to increase workflow management system scalability and performance to levels reached by typical transaction processing monitors.

3.6 Integration/interoperability among workflow management systems

The WfMC was formed in 1993 to promote interoperability among workflow management systems. Their standards address the areas of (i) APIs for consistent access to workflow management system services/functions; (ii) specifications for formats and protocols between workflow management systems themselves, and between workflow management systems and applications; and (iii) workflow model interchange specifications to allow the interchange of workflow specification among multiple workflow management systems. However, the current WfMC activities do not include more advanced types of interoperability such as those addressed by CORBA, X/Open, and the Web. Additional research is needed to decide appropriate workflow management system interoperability standards and advise the WfMC.

3.7 Interoperability/integration between workflow management system and heterogeneous information systems and applications

For workflow processes that access heterogeneous information systems, interoperability among heterogeneous systems and workflow management systems is important for the following reasons:

1. generic code allows access to heterogeneous information systems without recoding when these systems change,

2. the absence of code for routing and other forms of coordination enables fast development of applications with fewer errors, as opposed to applications that are developed using more conventional methods such as 4GL programming, and

3. it requires minimal workflow re-implementation to cope with changes in heterogeneous system functionality, since it requires no code changes in workflow implementations except re-specification of heterogeneous system interfaces.

Currently, interoperability means that various interface standards on different levels are available, such as protocol standards (e.g., e-Mail, TCP/IP), platform standards (e.g., MS-Windows, UNIX, Windows NT), and object interface standards (e.g., OLE, CORBA). However, providing further interoperability requires research to determine how to exploit and extend technology that complies with industry standards for interoperability, such as those developed by the Object Management Group and the World Wide Web Consortium. Because many types of errors could arise in a distributed heterogeneous computing environment in which the workflow is executed or enacted, error handling is generally considered to be a difficult problem. The difficulty is enhanced by the inherent complexity of organizational processes. Error prevention and handling is one research theme where a breakthrough is needed in order to deliver genuinely robust workflow processes.

3.8 Concurrency control, recovery, and advanced transactions

Issues of concurrency control are well-understood issues in database and transaction processing research and commercial products. State-of-the-art workflow management systems take different approaches to concurrency control as compared to database and transaction processing products, depending on perceived workflow requirements. Current approaches (e.g., check-in/check-out, pass-by-reference/pass-by-value, etc.) are primitive when compared to how database management systems support concurrency. Some workflow management systems allow multiple users/applications to retrieve the same data object concurrently. However, if each user decides to update that data object, new versions of the data item are created to be reconciled (merged) by human intervention. The rationale for this approach is the assumption that data object updates are rare. Thus, consistency can be handled by humans who review the data object versions and decide which one they want to keep.

To support limited forward recovery, contemporary workflow management systems utilize transaction mechanisms that are provided by database management systems. In particular, some workflow management systems use a database management system to maintain the state of the workflow instances, and issue database transactions to record state changes. In the event of a failure and restart, the workflow management system accesses the database management system to determine the state of each interrupted workflow instance, and attempts to continue executing workflow processes from the point they have been interrupted by a failure. However, such forward recovery is limited to the internal components of the workflow management system. State-of-the-art workflow management systems currently offer virtually no support for automatic undoing of incomplete workflow instances. Workflow designers may specify the withdrawal of a specific case from the system while it is running, possibly at various locations, for which an undo operation is needed at the process level (as opposed to the transaction level). When some organizational processes fail they can be compensated rather than rolled back. For example consider a workflow for purchasing a house. If the sale of a house is cancelled halfway, compensation payments must be made. Recovering from such failures requires that much of the recovery is designed specifically for this workflow application. Also, erroneous execution typically requires some form of human intervention. These issues illustrate that error detection, handling, and recovery are more complicated on the scale of an organizational process than they were on the scale of database transactions.

The research community is debating whether it is possible to use database management system technology and transaction processing monitor technology, or the extended/relaxed transaction models that have been developed to deal with the limitations of database transactions. New research is needed to determine viability for: (i) developing extended and application-specific transaction models for workflow application, (ii) integrating workflow specification with specification of transactional workflow properties, and (iii) developing workflow management system architectures that include and promote the use of extended transaction services.

3.9 Dataflow

In practice, the ability to pass data among the participants is what determines the effectiveness of a workflow management system. For example, during the processing of an international patent, each patent application involves a significant number of documents referencing other information sources and patents. Each existing patent is also a large collection of documents. Additional attached information such as articles and scientific papers are smaller in size but can be numerous. To the application itself, related patents and relevant articles are included as part of the documentation. Hence, transferring the case from the initial stages to the evaluation stages involves moving a great deal of information around (often from country to country as is the case with the European Patent Office).

The typical support provided for data flow is to ensure the existence of all information objects before an activity is started, and to locate and retrieve these objects. This typically requires no specific action on the part of the user, who will experience that all activities on the to do list come with all documents and information needed to do the work. Current workflow management systems achieve this by integrating an imaging system and the existing information systems with the workflow enactment service element or workflow server. However, such integration is often poor or application specific, and the server has minimal control over the flow of data, complicating the activity of adjusting it to the flow of control. Research is required to improve integration of imaging and dataflow, to deal with changing formats or structure of data as it makes way through a workflow process, modifications to data as well as to investigate alternative solutions that provide dataflow by moving references to data rather than the data itself, e.g., by using Databases, a CORBA ORB, or the Web.

3.10 Mobile clients and servers

Advanced workflow management systems allow mobile clients, i.e., workers and/or managers. More commonly, however, support for mobility is limited to distinguishing between failed and disconnected mobile clients. To receive work items, deliver results, and interact, users are required to periodically connect to a specific server of the workflow management system.

Providing support for mobile clients and servers is currently an open research problem in client/server architectures. Although the basic problem is not unique in workflow, the effects of mobility in workflow management involves workflow-specific research. Problems include: (i) workflow specification of mobile tasks involving applications and human performers, (ii) specification and management of migrating workflow processes, and (iii) tools for analysis and monitoring of mobile activities, agents, and workflow processes.

3.11 Web-based workflow systems and tools

The advent of the Web has made many workflow product designers consider Web-browsers as workflow client software. Using the Web as a front-end platform allows for workflow processes that are geographically spread out. Since many users already use Web-browsers, there is no need to distribute client software, thus enabling a wider class of applications. The robustness of Web-based workflow systems is a research problem, because the dialogue between Web clients and Web servers is based on a given language (HTML), which was not designed for this purpose. However, the prospect of inter-organizational workflow processes has come one step closer, and the research community is just beginning to understand the implications.

4 Future Directions

In the previous section we reviewed the current state of workflow and process management technologies and identified their main shortcomings. Some of these shortcomings are being addressed by the development efforts carried out by industrial vendors, while other constitute open questions that need to be addressed by research. In this section, we first review major areas of current research and then list problems that would need to be addressed in the future.

4.1 Current Research Areas

Overall, the focus of many of the current research efforts in the workflow and process management is the science and technological engineering of complex organizational processes. These efforts can be summarized in terms of the activities they address, which therefore represent a reasonable point of departure for highlighting the future directions for research in this area. Many of the tools, techniques, and concepts employed in current approaches to workflow and process automation studies, as characterized in earlier sections of this report, lay the foundation for the emerging research topics identified next.

- Definition and Modeling: Eliciting and capturing of informal workflow/process descriptions, and their conversion into formal workflow/process models or model instances that conform to their meta-model.

- Representation, Language, and Meta-Modeling: Constructing and refining a process/workflow concept vocabulary and/or ontology for representing families of organizational capabilities, language structures, use of logic/constraints/dependencies, etc. to specify activity coordination needs in a workflow process, and their instances.

  This enables the definition of the products, artifacts, roles, tools, and people as *resources* with interfaces, state (status), and other attributes that can be interconnected to other resources to satisfy process or workflow requirements at hand. Representations or meta-models must be designed to accommodate distributed multi-agent (w/multi-task agendas) problem-solving workflows or processes. Finally, we also observe that the representation, language, or meta-models employed anticipates (hence delimits) the kinds of real-world entities, conditions, situations, and patterns that can be specified. Thus, we must continually assess what must be specified to achieve some foundational, social, or technical purpose.

- Analysis, Testing, Verification and Evaluation: At least five different kinds of workflow and process analysis have been identified. This may suggest this is an area of growing research interest, both in terms of the kinds of analysis possible, as well as foreshadowing the kinds of tools, techniques, and concepts needed for workflow or process analysis. Development of useful matrices is an important goal.

  - *Logical*: Evaluating static and dynamic properties of a workflow/process model, including its consistency, completeness, internal correctness, traceability, as well as other semantic checks.
  - *Statistical*: Descriptive and inferential statistics which calculate the frequency, distribution, and association of selections, events, transactions, etc.
  - *Reasoning*: Query-based pattern-matching and inference to reason about space, time, organization (i.e., who, what, where, when, why, how, what-if, plus relational operators and filters), classification (taxonomic, genericity), configuration (composition, scheduling, replanning, generalization, specialization), and diagnosis (articulation).
  - *Feasibility*: Determining whether a proposed architecture can satisfy existing requirements, given available resources.
  - *Optimization*: Determining how to transform process flow to reduce resource utilization (e.g., reduce cycle time and cost).

- Simulation: Simulation of complex processes and workflows provides an important source of insight into the behavioral dynamics of the simulated work activity. Accordingly, simulation tools, techniques, and concepts can be used as an aid in: (a) refining process or workflow performance requirements, (b) rehearsal and training of new users (e.g., as organizational flight simulators), (c) prototyping and validating alternative workflow or process designs, and (d) for other forms of workflow design support. As such, at least three kinds of simulation technologies are being investigated in the field.

  - *Knowledge-Based Simulation*: Symbolically enacting process models in order to determine the path and flow of intermediate state transitions in ways that can be made persistent, replayed, queried, dynamically analyzed, and reconfigured into multiple (what-if) alternative scenarios.
  - *Discrete-Event Simulation*: Computationally enacting a sample of process models as network flows with heuristic or statistical arrival rates and service times so as to determine the overall process performance envelope, throughput, systematic behavior, and resource bottlenecks.
  - *Continuous Flow Simulation*: Modeling a complex process in a form that represents a continuous system. Continuous systems can in turn often be expressed as relationships among a system of linear and non-linear equations, whose behavior can be derived by solving the equations.

- Prototyping, Walk-through, and Performance Support (Training On Demand): Incrementally enacting partially specified workflow/process model instances in order to evaluate process presentation scenarios to end users, prior to performing tool and information system integration. Automatically deriving or generating a web of views of the modeled workflow/process in the form of a performance support or interactive, on-demand training system tailored for the process or workflow at hand.

- Administration, Staffing and Scheduling: Assigning, scheduling, and binding specified users, tools, and development data objects to user roles, product milestones, and work schedule defined for an instance of a workflow/process model. Tracking and reacting to workloads on human participants in a workflow/process.

- Interoperation and Integration: Data, tool, user interface, workflow/process providing a client-side view or ***workspace*** as a computational context at the granularity of input and output resources. Such a context denotes a cached web binding user, org-role, task, tool(s), and enactable actions/steps. Subsequently, this entails encapsulating or wrapping selected information systems, repositories, data objects that are to be invoked or manipulated when enacting a workflow/process instance, interoperation among different workflow/process designers, and run-time (enactment) systems. Exploiting latest advances in infrastructure technologies, including distributed computing (as supported by, for example, distributed transaction monitors and CORBA), collaboration, distributed database and Internet technologies is an important concern.

- Target Support Environment Generation: Automatically transforming a process model or instance into an executable code for the target (possibly distributed and heterogeneous) computing environment that presents itself to end-users for process enactment, or develops a workflow application ready for invocation.
- Monitoring and Measurement: Collecting and measuring process enactment data needed to improve subsequent process enactment iterations, as well as documenting what process actions actually occurred in what order.
- Visualization: Providing users with graphic views of process model definitions, simulated/executed workflow instances, and actual executions, that can be viewed, navigationally traversed, interactively edited, and animated to convey statics and dynamics of process workflows. 3D visualizations software and systems will likely be utilized.
- Enactment History Capture and Replay: Interactively capturing and graphically visualizing the re-enactment of a process, in order to more readily observe process state transitions or to intuitively detect possible process enactment anomalies.
- Fault Detection, Error Handling or Repair: Diagnosing, replanning, and rescheduling actual or simulated process (instance) enactments that have unexpectedly broken down due to some unmet workflow resource requirement, contention, availability, or other resource failure. This can require a diagnostic classification taxonomy, heuristic/algorithmic repair taxonomy, and workflow/process instance update mechanism, ability to detect errors, and procedures to handle them.
- Evolution, Continuous Improvement and Model Management: Evaluation, incremental and iterative enhancement, restructuring, tuning, migrating, or reengineering workflow/process models – and workflow/process life cycle activities – to more effectively meet emerging user requirements, and to capitalize on opportunistic benefits associated with new tools, techniques, and concepts.

4.2 Future Research Directions

In addition to these activities and observations which are currently receiving varying degrees of attention within the field, there were a number of other topics that workshop participants indicated were receiving too little attention, or none at all.

- Reference baselines and empirical case studies
  - *challenge problems*: how would the community and industry be better served if there were a sample of challenge or 'grand challenge' problems were workflow and process automation are central components?
  - *comparative benchmarking*: how could reference architectures, performance assessments, and modest reference problems be employed to defined research or industrial benchmarks?
  - *comparative case studies*: how should empirical studies of workflow and process automation tools, techniques, concepts, and their use by different people in various organizational settings be structured so that the results can be readily shared (e.g., in open WWW-based repositories), compared, and extended?
  - *baseline cases and stress testing cases*: how can challenge problems, benchmarks, and case studies be structured to serve as baselines or stress testing (i.e., pushing to the effective limits)?
- Representational concepts, language constructs, or other notational devices
  - What types of concepts, constructs, or notational features should be considered as constituting the baseline for specifying a complex organizational process or workflow?
    * Objects, attributes, values, relations, control flow (sequence, iteration, concurrency, conditional), computational methods and rules, and other operations.

- What types of representational extensions need further investigation and experimentation?
- Given a representational medium in which complex processes and workflows can be specified (formally or informally), a variety of views or *centering perspectives* may be needed. These perspectives serve to reveal or highlight foundational, technical, or social aspects of workflows and processes that have been specified, or that are automated. These perspectives include, human/group, work, process, tools, artifact/product, design, planning, etc.
- Finally, we need to identify some of the purposes for which processes and workflows are specified and executed, so that we can assess how well the modeled processes and executable workflows meet various social and technical needs. These include, communication, understanding, execution, fidelity, and analysis.

- What research approaches/methods are needed? Beyond the activities and observations outlined above, there remains an additional set of topics that merit further investigation, but do not fall cleanly into the classification scheme employed above. As such, these topics can be listed as follows, without any qualification as to their significance in comparison to those items already described.
  - We need a *methodology for transition of technology* from researchers to industry or users and from industry or users to researchers.
  - How to support process or workflow interoperability?
  - New organizational forms enabled by workflow and process automation
  - Representing reflections and reflexive relationships
  - Process reuse

5 Directions for Community Building

As the workshop helped establish the roots of workflow and process management in multiple disciplines, an important proposed activity is the creation of the community that will promote the development of this emerging field of work activity coordination. This community should consist of scientists and researchers in academia and industrial laboratories, software developers, workflow product vendors, application vendors, and users who can contribute to the articulation of requirements for new software.

Web technologies can play an important role in the emergence and maturing of this community, serving as a medium for exchange of information and reference resources. Some of the useful kinds of information that may be disseminated via a Web-based repository of reference material and resources may include:

- technical reports and pointers to other bibliographic material
- demos of prototype systems and applications
- project lists and description
- glossary of terms
- proof-of-concept software tools (offered as freeware to other researchers)
- offers from commercial vendors to promote research by providing free or significantly discounted infrastructure technology and tools, such as
  - relational database systems
  - (object oriented) database systems
  - multidatabase systems
  - object brokers (multi-platform, multilanguage)
  - process engines
  - GUI-builders
  - transaction monitors

An important step in the formation of a workflow community could be a representative workshop or conference where all constituent (sub-) communities are represented (software process, databases, computer information systems, organizational process modeling, users). In particular, the workshop should be attended by participants selected by the respective communities. In addition to research papers, such a workshop or conference should provide:

- tutorials, representing the state of the art in various sub-disciplines,
- focused presentations and demonstrations of software and systems,
- panels and discussion groups.

A parallel step can be a conscious effort to increase the cross-fertilization of ideas by inviting members of complementary communities to participate in each other's workshops and conferences. Such meetings could be used to generate a common set of standard (baseline) and challenge problems. It should be recognized that while standardization efforts will be led by trade organizations (such as WfMC), the rest of the community should participate in these efforts.

Finally, it should be emphasized that research funding organizations should play a role in creation and support of various community building activities, including workflow-related workshops and/or conference. An inter-agency program including NSF, DARPA, and ESPRIT would go a long way in creating impetus for workflow activities. However, the long term viability of the discipline will ultimately require that connections between the academia, industrial labs and industry itself are created to generate sufficient support for research, in this practical discipline, with significant engineering component.

Bibliography

The BIS Guide to Workflow Software: A visual comparison of today's leading products. B. R. Silver. BIS Strategic Decisions, Norwell, MA, September 1995.

Business Process Change: Reengineering Concepts, Methods and Technologies. V. Grover and W. J. Kettinger. Idea Group Publishings, 1995.

Proceedings of the First International Conference on the Software Process. IEEE Computer Society Press, October 1991.

Proceedings of the Second International Conference on the Software Process. IEEE Computer Society Press, February 1993.

Proceedings of the Third International Conference on the Software Process. IEEE Computer Society Press, October 1994.

Proceedings of the NSF Workshop on Workflow and Process Automation in Information Systems: State-of-the-art and Future Directions . A. Sheth, Editor. The University of Georgia, May 1996.

Process-centered Software Engineering Environments. P. K. Garg and M. Jazayeri, Editors. IEEE Computer Society Process, 1996.

Software Process. A. Fuggetta and A. L. Wolf, Editors. *Trends in Software.* No. 4, Wiley, London, 1996.

Software Process Modeling and Technology. A. Finkelstein, J. Kramer and B. Nuseibeh, Editors. Wiley, London, 1994.

Special issue: Groupware for self-organizing units. B. Lange. In *SIGOIS Bulletin.* ACM Press, New York, NY, April 1996.

Special issue on software support for workflow management. A. Bukhres and E. Kuhn, Guest editors. *An International Journal of Distributed and Parallel Databases.* Kluwer Academic Publishers, Hingham, MA, April 1995.

Special issue on workflow and extended transaction systems. M. Hsu, Editor. *Bulletin of the IEEE Technical Committee on Data Engineering.* IEEE Computer Society, Washington, DC, June 1993.

Special issue on workflow systems. M. Hsu, Editor. In *Bulletin of the Technical Committee on Data Engineering.* IEEE Computer Society, Washington, DC, March 1995.

Special-theme issue: Multidatabases. A. Dogac, Editor. In *Journal of Database Management.* Idea Group Publishing, Harrisburg, PA, Winter 1996.

WA-12: An Empirical Study about the Practice of Workflow Management. S. Joosten, G. Aussems, M. Duitshof, R. Huffmeijer, and E. Mulder. Research Monograph. University of Twente, enschede, The Netherlands, July 1994.

The Workflow Management Coalition. Glossary – a workflow management coalition specification. The Workflow Management Coalition, Brussels, Belgium, November 1994.

The Workflow Paradigm - The Impact of Information Technology on Business Process Reengineering (Second Edition). L. Fischer. Future Strategies, Inc., Alameda, CA, 1995.