A. Project Summary

A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries

Problem Statement: Digital libraries face significant challenges beyond providing traditional library services (such as information retrieval, indexing resources, and archiving materials). New challenges include supporting the collaborative work of a digital library's stakeholders, providing analysis tools, and hosting simulated or virtual environments.

A particularly difficult problem associated with the first of these challenges—supporting collaborative work—is managing the workflow associated with digital library activities, such as reviewing a new submission or classifying existing materials. Digital libraries have unique workflow requirements because their workflows typically involve multiple stakeholders distributed across different organizations who must coordinate their actions.

Complicating matters, a particular activity, such as reviewing new materials for inclusion in the digital library, may involve a different subset of stakeholders each time it is initiated. For instance, a digital library specializing in literature, may send materials for 19th century literature to university A for review, while materials for 18th century literature are reviewed by university B. Thus, the same activity (reviewing new materials) involves university A in one instance and university B in another.

Further complicating matters is that for each activity, each stakeholder may have a different set of steps they use to complete the activity. Thus, university A may require two faculty members to review and approve new materials, while university B only requires a graduate student in the appropriate area to approve the new submission.

Finally, because so many different stakeholders are involved, traditional approaches to workflow systems are often inappropriate. For instance, a monolithic workflow system is ill-equipped to manage processes across multiple organizations while distributed workflow systems typically assume that they are being deployed in the same organization, where problems of "who owns the process" are minimized. In addition, workflow systems sometimes assume a common computing platform will be used across stakeholders, which is an inappropriate assumption in the context of a digital library.

Approach. Recent advances in event messaging systems, hypermedia, and Web-based technologies (such as XML), provide new opportunities for creating lightweight workflow technology that can help the stakeholders of a digital library coordinate their shared activities. We propose to conduct research into developing new infrastructure for deploying workflow services into digital libraries that can be tailored to meet their unique needs. In particular, we intend to focus on new abstractions for workflow processes that can handle the heterogeneity of digital library activities (e.g. distribution across multiple organizations, instantiation of the same activity across different subsets of digital library stakeholders, each with their own way of performing the assigned tasks, etc.) and new techniques for making workflow technology "lightweight." Lightweight technology should have a low entry barrier to use, support multiple computing platforms, and provide a certain level of autonomy to the individual organizations that adopt it.

In particular, we intend to develop technology that will allow multiple organizations to create and share workflows. Workflows will be specified in an XML-based format and can be centralized on a common workflow server for a digital library, or distributed across the servers of multiple stakeholders. Tools will be provided to create, manipulate, and link workflows across multiple stakeholders. Event notification technology will be used to capture significant workflow events and notify all stakeholders dependent on those events. In addition to using XML, Web-based technology, such as Java servlets, will be used to lower the entry barrier to use for the proposed workflow system and ensure that all workflow services are accessible by widely deployed Web browsers.

Evaluation. In order to evaluate our techniques and technology, we intend to model and support the workflow processes of three "real-world" digital libraries. The first digital library is the Web Portal to the Scientific Computing Division of the National Center for Atmospheric Research. This portal must provide its users access to SCD's formidable computing environment, manage the collaborative creation of scientific documents, and index and archive scientific resources. Our second evaluation target is the JESSE electronic journal, associated with the Digital Library for Earth System Education, or DLESE, which is a digital library that is dedicated to the collection, enhancement, and distribution of materials that facilitate learning about the Earth system at all educational levels. Our third evaluation target is the JIME electronic journal. JESSE and JIME, in particular, experience some of the workflow problems characterized above, and should provide excellent testbeds for evaluating the utility of our new workflow abstractions and our ability to create workflow technology that is easy to adopt and integrate into an existing digital library.

C. Project Description

A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries

1 Introduction

Digital libraries face significant challenges beyond providing traditional library services (such as information retrieval, indexing resources, and archiving materials). One key challenge involves supporting the collaborative work of a digital library's stakeholders. In particular, it is challenging to coordinate the actions of these stakeholders as they work together to accomplish various activities, such as reviewing a new submission or classifying existing materials. The difficulty lies in the fact that the stakeholders of a digital library—librarians, reviewers, authors, students, etc.—are highly distributed and diverse, and do not necessarily share working styles, goals, or computing platforms.

Complicating matters, a particular activity—such as reviewing new materials for inclusion in the digital library—may involve a different subset of stakeholders each time it is initiated. For instance, a digital library specializing in literature, may send materials for 19th century literature to university A for review, while materials for 18th century literature are reviewed by university B. Thus, the same activity (reviewing new materials) involves university A in one instance and university B in another. Further complicating matters is that for each activity, each stakeholder may have a different set of steps they use to complete the activity. Thus, university A may require two faculty members to review and approve the new materials, while university B only requires a graduate student in the appropriate area to approve the new submission. These situations lend insight into why it is so difficult to effectively coordinate the activities of a digital library's stakeholders: Each instance of a digital library activity is a highly specialized artifact. This specialization makes it particularly difficult to track and compare the progress of particular activities. Again, two instances of the "review new materials" process may both be at the same step in the process, but one may complete much more rapidly than the other based on the performance of the organizations that have responsibility for that particular step.

Traditionally, coordinating a group's activities has been accomplished using software process environments or workflow systems [3, 6]. However, many traditional approaches to this technology make assumptions about their users that are inappropriate within the digital library context. For instance, some workflow systems assume that all users work for a single organization, or even worse, assume that all users are members of the same team within a single organization [9]. Other workflow systems are only slightly better in that they can support multiple users across multiple groups, but once again make the assumption that all users work for the same organization [9]. These assumptions make certain aspects of the workflow problem easier to manage. For instance, members of the same organization often have access to similar computing platforms, and problems such as "who owns the process" are minimized, often because responsibility is dictated by organizational structure. However, these assumptions are simply invalid in the digital library context, where stakeholders come from different organizations (or have no affiliations at all, e.g. a student or customer of the digital library) and are distributed both in space and time. Therefore, new research is needed to accommodate the unique workflow requirements of digital library activities.

We propose to conduct research in this area with the goal of developing new workflow techniques and tools for digital libraries with the following characteristics:

1. **Lightweight.** Lightweight workflow technology should have a low entry barrier to use, support multiple computing platforms, and provide a certain level of autonomy to the organizations that adopt it.
2. **Flexible.** Workflow technology needs to conform to the organizations and individuals that use it, not the other way around. As such, workflow technology needs to be flexible and customizable. We intend to build flexibility into our techniques and tools by basing our workflow abstractions around *events*. We believe it will be straightforward to construct event-based workflows that are easy to manipulate and evolve.
3. **Web-based.** The World Wide Web [7] has had tremendous impact in deploying information technology pervasively across multiple organizations. Indeed, with nearly 30 million web servers detected in the February 2001 Netcraft Web Server Survey (up from 3.1 million in September 1998 [4]), it is becoming possible to assume that individuals in most organizations have access to Web browsers, if not Web servers.[1] As such, workflow technology needs to be accessible via Web-based services and it must be possible to share and link workflows between multiple organizations via the Web.

1. The Netcraft Web Server Survey is available at <http://www.netcraft.com/survey/>.

We provide details on how we plan to achieve these characteristics in Section 3

In order to evaluate our techniques and technology, we intend to model and support the workflow processes of three "real-world" digital libraries. The first digital library is the Web Portal to the Scientific Computing Division (SCD) of the National Center for Atmospheric Research (NCAR). This portal must provide its users access to SCD's formidable computing environment, manage the collaborative creation of scientific documents, and index and archive scientific resources. Our second evaluation target is the JESSE electronic journal, associated with the Digital Library for Earth System Education, or DLESE, which is a digital library that is dedicated to the collection, enhancement, and distribution of materials that facilitate learning about the Earth system at all educational levels. Our third evaluation target is the JIME electronic journal. JESSE and JIME experience some of the workflow problems characterized above, and should provide excellent testbeds for evaluating the utility of our new workflow abstractions and our ability to create workflow technology that is easy to adopt and integrate into an existing digital library.

To support these efforts, we have assembled a team of researchers and practitioners qualified to perform the research agenda proposed here. The principal investigator (PI) from the University of Colorado, Boulder has a successful record of research in software engineering, hypermedia, and software architecture particularly in addressing issues of scale, heterogeneity, and event notification which are all critical aspects of producing Web-based workflow technology that is lightweight and flexible. The Co-PI is the editor of the JESSE electronic journal and thus has deep insight into the workflow problems confronting JESSE's stakeholders. The senior person from NCAR is the Operations Section Manager of the Scientific Computing Division and is thus intimately familiar with SCD's Web Portal as well as broadly knowledgeable about the patterns of use, or workflows, encountered in this environment. Finally, while the editors of the JIME journal are not directly involved in the proposed research effort, they have written a letter of support for the proposed research. See the Supplemental Documentation section of the proposal for more information.

This project will result in the following advancements to digital library infrastructure:

1. An understanding of the characteristics of digital library workflows and the types of services required to support and coordinate them.
2. An understanding of the heterogeneity encountered in digital library work environments and the techniques required to create workflow technology that can effectively handle this heterogeneity while providing a low entry barrier to use.
3. An understanding of how improved support for workflow can impact the ability of a digital library to meet the needs of its users. It is our hypothesis that automated support for workflow can make most digital library activities faster and more efficient. Such a benefit would provide the stakeholders of a digital library with more time to focus on other aspects of a digital library, such as the quality of its materials or more timely updates, that will increase the overall utility of the library for its end-users.

In addition, the proposed research should provide benefits to higher education in the form of new educational materials for computer science courses at the undergraduate and graduate levels. In particular, new instruction, course work, and project ideas can be based on the increased understanding of the nature of workflow in digital library environments. These new materials can be applied to existing courses on software engineering (the traditional domain where workflow is taught) or to new courses in hypermedia and digital libraries.

The rest of this proposal is organized as follows. Section 2 presents information on the types of workflows that exist in JESSE and NCAR's Web portal. Section 3 describes the approach of the research in more detail and outlines a specific research plan for the two years of the project. Section 4 reviews related work. Section 5 presents the potential impact of the proposed research on higher education. Section 6 briefly discusses results from prior NSF-funded research. Finally, Section 7 concludes the project description with a summary of the project's research goals and deliverables.

2 Example Digital Library Workflows

We motivate the need for new research to support digital library workflows by presenting examples drawn from the JESSE electronic journal and NCAR's Web portal. While these systems are not digital libraries in the strict sense of the definition, they posses many of the same characteristics and confront similar activities and workflows. As such, workflow technology developed to support the issues and workflows described below, will be directly applicable in digital library contexts as well. After we present the examples, we conclude this section with a discussion of the requirements these scenarios place on workflow technology for digital libraries.

2.1 The JESSE Review Process and Timelines

The Journal of Earth System Science Education (JESSE) is a new interdisciplinary journal aiming to foster the study of the Earth as a system and promote the development and exchange of interdisciplinary learning resources for formal and informal education. JESSE will publish a wide ranging variety of electronic content, with minimal constraints on format, targeting undergraduate educators and students as the principal readership, expanding to a middle and high school audience as the journal matures. JESSE aims for rapid review and turn-around of resources to be published, with a goal of 12 weeks from submission to publication for resources requiring few changes. JESSE employs an open peer review process in which authors and reviewers discuss directly the acceptability of a resource for publication using a software tool called the Digital Document Discourse Environment. Reviewer comments and attribution will be available with the resource upon acceptance for publication. In the development phase, JESSE will also conduct a parallel anonymous review of content to validate and ensure credibility of the open review approach. JESSE is collaborating with the Digital Library for Earth System Education (DLESE) as a federated partner.

Unlike traditional publication, JESSE will be reviewing contributed materials that are in many cases already being made available via the Internet. JESSE will review and publish only those resources that have been submitted to the Journal by the author(s). Contributions may be solicited or unsolicited. JESSE or its partners may solicit authors directly to contribute specific materials that are already in use or being distributed on the Internet. As the Journal matures the number of unsolicited contributions from authors is expected to increase.

JESSE will be updated on a continuing basis with new beta resources posted as they become available, and resources published as reviews are completed. As such, the journal has no fixed publication dates. For the purposes of review however, orderly timelines will be established to assure timely publication of submittals. What follows is a description of the path for a learning resource through the JESSE peer review and publication process.

1) (Week -n) Dr. X of ABC University has developed material exploring phytoplankton and coastal ocean productivity using satellite observations. Dr. X's material is designed to be completed in 6 hours of undergraduate laboratory time. The exercise is relatively self contained (except for links to current satellite data) and resides upon ABCU's web server as a collection of hypertext, images and references for Dr. X's laboratory use.

2) (Week 0) The JESSE editors, seeking to populate the archive with a range of interdisciplinary resources, invite Dr. X to submit the phytoplankton module to JESSE for peer review and publication. Dr. X agrees. Alternatively, Dr. X can submit the module directly to JESSE, unsolicited, or one of the JESSE partners can solicit Dr. X's contribution.

3) (Week 1) Dr. X accesses the JESSE web site and reviews the procedures and guidelines for submitting materials to JESSE. The submission process is simplified through the use of an electronic template. The web interface collects background information about the resource, including authorship, intended audience, content and learning level, module context, etc., and a hyperlink to the resource itself on the ABCU server.

4) (Week 2) Dr. X's information about the module is added to the JESSE beta resource database as a learning resource awaiting review. Educators searching the beta area will be able to view the supplemental background information and volunteer or recommend others to serve as beta testers/reviewers of the module.

5) (Week 2) While browsing the JESSE site, Dr. Y identifies the phytoplankton module as useful and volunteers to try the module in the classroom and serve as a potential reviewer. The JESSE web page provides Dr. Y with review criteria and guidelines. The editorial staff records Dr. Y's interest and registers him/her as a reviewer.

6) (Week 3) JESSE will maintain a database to track submissions, authors, and reviewers in managing the flow of learning resources through the review process. This database is accessible on a private web site available only to the editorial staff. The Managing Editor works with the Principal Editors to designate a CoEditor who is responsible for identifying reviewers in addition to Dr. Y who are qualified to assess the content, pedagogy, and presentation style of the phytoplankton module. The CoEditor is responsible for shepherding the process and mediating the discussion between reviewers and author. The CoEditor, reviewers, and author establish a timeline for the review based upon a standard template. A review completion date is designated.

The JESSE editorial staff will prepare the learning resource documentation for the Digital Document Discourse Environment (D3E) in preparation for the review process. D3E and email provide the basic mechanisms for communication and documentation during the review process.

7) (Weeks 3-9) The reviewers work with Dr. X directly to answer any questions about the module, and provide recommendations for improvement and ideas for use in the classroom and laboratory via D3E. The reviewers' interaction with Dr. X is confidential. When the author and reviewers reach agreement upon the acceptability of the module, the CoEditor summarizes the review comments and suggestions. This information is archived with the module, along with final reviewer commentary, and becomes a part of the background commentary for the resource.

The open reviewer/author relationship is designed to enable non-adversarial constructive recommendations and the exchange of commentary [21]. The CoEditor ascertains that the review criteria are properly applied and in retaining the final decision concerning publication assures the merit of the submission. If major changes are required, the CoEditor and author will agree upon a timeline for re-submission, or the author may choose to withdraw the submission. In either case reviewer comments and attribution are maintained confidentially for a limited time.

8) (Weeks 10-12) Dr. X's phytoplankton module is accepted for publication. Additional information about the module such as the synopsized reviewer comments and Dr. Y's classroom recommendations are archived in a public area of the D3E. The module contents are stored and maintained as accepted on the JESSE server, and held under configuration control (i.e. change protection) as an archive of the reviewed resource.

The public area of the D3E which organizes accepted learning resources becomes the focal point for the peer commentary function of JESSE. This function will enable the community at large to comment on the usefulness or applicability of the resource, which however will be moderated to protect the interests of all, especially the authors.

9) Subsequent to week 12 classroom educators download the phytoplankton module assured of its quality and guided in its use with reviewer and peer commentary while Dr. X adds a peer reviewed publication to his/her vita.

2.2 Workflows associated with the NCAR Web Portal

NCAR provides services and facilities to support a broad, geographically distributed and loosely federated user community made up of 66 member universities and numerous other collaborators. NCAR has a duel yet complementary purpose of furthering research fundamental to the understanding of the earth system, particularly the atmosphere, and its interactions from the land and ocean surface up to the Sun. Also fundamental to NCAR's mission is making the research results available to educators in K-12, Undergraduate and Graduate programs. Because of these two broad roles and NCAR's unique facilities such as aircraft, radar and supercomputers, NCAR is very interested in providing better tools to facilitate the collaborations within this community that is geographically distributed across diverse organizations. Below, we present three examples of typical NCAR workflows and identify how the proposed workflow technology can be used to augment the ability of NCAR's scientists to complete these activities successfully.

1)Dr. X along with a team at NCAR is deployed on a field campaign on the great plains region of the western United States. They are utilizing an array of equipment including specialized Doppler radar to research and study severe thunderstorms including those that have a high probability of generating tornados. The team of researchers in the field along with the team back at NCAR can more effectively share information by collaboratively constructing the workflow they will use to gather, package, and analyze the collected data. Because their workflow is accessible via the NCAR Web Portal, the scientists in the field are able to update the workflow based on changing weather conditions and their colleagues back at NCAR are notified of these changes automatically by the shared workflow system.

2)The data collected by Dr. X's team is sent to both NCAR and the Severe Storms laboratory at the University of Oklahoma for analysis. Dr. Y at NCAR is interested in comparing the observed data against the results generated by her new computer model, while the Severe Storms lab is interested in isolating the factors that lead to the strongest and most dangerous type of tornados. Both Dr. Y and the Severe Storms lab submit job requests to the Web Portal to access the data being streamed by Dr. X's team. Since job requests are modelled as workflows (e.g. first invoke tool A, then invoke tool B, then invoke tool C, etc.), the workflow system processes the requests and detects that the same data set is being requested. A meta-workflow causes an analysis program to be invoked and it is further determined that the University of Oklahoma and Dr. Y are interested in different portions of the data stream. This information allows optimizations to be applied to the delivery of the data and both parties are able to perform their experiments without unduly impacting the other. As Dr. Y proceeds in her analysis, she constructs a workflow that allows her associates, grad students and programmers interested in the results of her analysis to be notified when new data is available.

3)Dr. Z at the University of Oklahoma elects to package some interesting sets of data streaming from the campaign into curriculum material for his graduate course in meteorology. Dr. Z submits this material to the Web Portal for internal NCAR review, using a workflow similar to the one described for JESSE. (In addition, NCAR will also need to

support review cycles similar to classical peer review mechanisms.) Once approved the material can be made available on the Web Portal to other educators teaching similar courses.

Currently the activities described above are performed manually, with ad hoc, individually compartmentalized, techniques for moving data, technical documentation, laboratory notes, etc. to the places they are needed. Current techniques include fax, e-mail and conference calls. Workflow tools could thus greatly enhance these activities and make it easier to manage, measure, and evolve the processes NCAR employs on a day-to-day basis. In addition, the use of workflow technology can provide value-added services such as automated meta-data creation and metric collection.

2.3 Requirements for Digital Library Workflow Technology

The scenarios above suggest certain requirements for workflow technology designed to support digital libraries. In particular, the concerns raised in the Introduction are reinforced:

1. Digital library activities involve multiple stakeholders from multiple organizations. In the JESSE scenario, stakeholders include the author, the managing editor, the principal editors, the coeditor, the reviewers, the educators, and the students. In the NCAR scenario, stakeholders include various scientists with potentially different affiliations, reviewers, "colleagues", and students. Workflow technology for digital libraries must be able to coordinate the actions of multiple stakeholders from multiple organizations, and allow for differences in the way particular activities are carried out by various stakeholders. For instance, each reviewer in the JESSE review process may conduct their reviews and interactions with the author in different ways. (Some may engage in constant dialogue with the author, for instance, while others operate in "batch mode" interacting with the author only at specific times in their review cycle.
2. Digital library stakeholders will not possess a uniform computing environment. The supercomputers of NCAR are not available for use by Dr. X's team in the field (who are presumably using laptop computers connected to the Internet via satellite). The author in the JESSE scenario may have very different computing resources than the reviewers who review his course and the students who eventually use his course materials. As such, workflow technology for digital libraries must be able to coordinate the actions of stakeholders across a heterogeneous set of computing platforms. As mentioned in the Introduction, a key characteristic to achieving this requirement is that the workflow technology should be lightweight.
3. Digital library stakeholders benefit from having their process accessible. For instance, Dr. X's team was located in a highly dynamic environment (thunderstorms) and needed to change their activities (data collection, storm pursuit, tornado avoidance, etc.) with some frequency. Having the ability to access and update their process or plan of work dynamically (and having these changes communicated to their colleagues) is critical to the successful coordination of the data collection activity. The requirement of accessibility can be met in many ways, for instance a cell phone interface or handheld computer interface could potentially serve equally well as a Web-based interface. We are proposing to make use of Web technologies to enable access to shared workflows to target a significant portion of digital library stakeholders. It is far more likely that all stakeholders of a digital library will have access to a Web browser than a cell phone or PDA, despite the increasing popularity of these devices.
4. Activities may require the invocation of particular tools at certain steps in the process. The second NCAR example, for instance, required that the workflow technology be intelligent about how the job requests of various stakeholders interacted with the data sets that exist within the digital library. This requirement is not unique to the digital library domain, and indeed, associating tools with particular steps of a workflow is a common feature of all workflow systems. The unique aspect that comes to play in the digital library domain is related to the previous two requirements in that the workflow technology cannot make the assumption that each stakeholder in the workflow has access to the same set of tools. Indeed, the scientists in the field cannot invoke applications on their laptop that are available to their colleagues back in NCAR's supercomputing environment. As such, workflow technology for digital libraries must be able to determine if alternative tools exist that can be invoked for a particular stakeholder's environment or if a Web-based application can be invoked to provide access to required functionality using a platform-independent mechanism.
5. Digital library activities are not necessarily sequentially organized. The JESSE timeline in particular describes an activity that is only loosely arranged in a sequential fashion. During weeks 3-9 while the review is being conducted, potential actions are highly parallel with reviewers accessing the materials and sending comments to the author, the author updating the materials in response, and the CoEditor facilitating discus-

sion and monitoring the quality of the reviewers comments. This type of workflow is very difficult to model using traditional workflow formalisms. We propose to base our workflow formalisms on events, and allow the specification of event-based workflows that can very flexibly handle highly parallel activities such as the JESSE workflow. Event-based workflows should also be effective in dealing with multiple instances of the JESSE workflow, which is a requirement since the JESSE electronic journal is in a continual mode of review with no preset publication schedules, unlike traditional paper journals.

Based on these requirements, we describe below our approach to developing workflow technology for digital libraries.

3 Approach and Research Plan

3.1 Approach

Our approach to workflow technology for digital libraries is predicated on the notions of flexibility, usability, and accessibility. Our workflows will be defined in terms of events and our technology will be based on an existing Internet-scale event messaging system (use of an existing technology will save time, allow us to focus on just the workflow aspects of the problem, and makes feasible the completion of the proposed research in the two year time frame). Event-based workflows provide flexibility because it is easy to add new events to a workflow as they become necessary and remove events that are no longer effective or needed to achieve the goal of the activity. Usability is enabled both by our requirement to keep the entry barrier to use as low as possible and by providing graphical notations for workflows that can be created, manipulated, shared, and linked using a Web browser. Finally, our use of Web based technologies provide anytime, anywhere accessibility to the workflow technology increasing the ability of the digital library stakeholders to coordinate their actions and achieve their goals. In this section, we describe in detail the approach we use to achieve these characteristics, our evaluation plans, and present a research plan that details the major milestones for the two year time frame of the proposed research.

3.1.1 Event-based Workflows

Workflow technology typically organizes activities into a set of steps that explicitly states what needs to be done first, what second, and so on until the end of the activity. Workflow models allow workflows to be organized sequentially or in parallel, with a variety of constructs that allow for conditional processing (e.g. if X, do A, else do B), forks (do A and B at the same time), and joins (do C only after A and B are finished). While this approach provides a lot of power, it encounters one major difficulty: it requires that a process be defined formally, but people rarely follow a defined process when performing work. The reason for this is that a worker's actions are highly dependent on the situation and context in which they are performed [22]. Thus, a worker may perform the same task differently each time they do it, because, for example, in one situation they may be time constrained (and thus will skip steps) whereas in another situation they may be bored (and thus make mistakes) and in another situation they may be training a new employee (and thus be extra thorough). In addition, events tend to be second-class citizens in these types of workflow systems, since they mainly correspond to announcing that a particular step is complete (e.g. step 1 is done, step 2 is done). The problem with this situation is that events can be a rich formalism for capturing domain knowledge and conveying information to workers that is semantically meaningful. Thus, "step 10 is complete" means nothing to a worker but "Code Inspection Complete" conveys much more information and may cause a worker to anticipate future work assignments. For instance, a reviewer who receives a "new submission" event and discovers that the topic of the new submission is in his/her area, can anticipate that they will be assigned to review the submission and can plan accordingly.

We propose to design workflow technology that constructs activities around semantically-meaningful, domain-specific events. Workflows are specified as a set of event patterns that describe the structure of an activity. We initially intend to base our specification of event patterns upon the notation of regular expressions. Thus, a pattern of the form:

```
NewSubmission, AutoReject?, EditorAssigned, ReviewerAssigned+, ReviewReceived+,
[ReviseAndResubmit | Accept, Publish | Reject]
```

describes a fairly typical review process of a traditional journal in terms of events. First a submission is received, which may be automatically rejected by the editor-in-chief if it is inappropriate for the journal. Otherwise an editor is assigned, who finds reviewers, waits for their reviews, and then either accepts and publishes the paper, rejects it, or recommends that it be revised and reviewed again. Note: that regular expressions are only a starting point, since they are not capable of some desired functionality. For instance, the pattern above states that at least one (but possibly more) ReviewerAssigned events must occur followed by at least one (but possible more) ReviewReceived events. However, regular expressions have no way of specifying that the number of ReviewerAssigned events should equal

the number of ReviewReceived events, but clearly if we assign three reviewers, we should wait until we have at least three reviews before proceeding!

Actions can be associated with each event, such that if an event occurs, the associated actions can be triggered, e.g. "if a 'new submission' event occurs, transfer the submission to a private FTP directory and send an e-mail to the editor-in-chief containing the cover letter provided with the submission." This capability is very powerful since it provides for the specification of complex workflows in a decentralized manner, e.g. at the same time that the editor-in-chief receives the "new submission" event, all of the associate editors receive it as well, and may receive meta-information about the submission such as the title, topic keywords, and length of submission. The associate editors can use that information to decide if they are interested in managing the review process for the submission or they can use it to anticipate that the editor-in-chief will be assigning the submission to them in the near future. Note: depending on the sophistication of the submission process, there may be enough information to generate a more specialized event, such as "new 18th Century Literature submission." In this case, only those editors interested in 18th century literature will receive the "new submission" event.

Web-based tools will be developed to allow the stakeholders of a digital library to define and share event types, define and share workflows by specifying event patterns, check the status of a particular activity by viewing the events that have occurred, and defining the actions that should be invoked when a particular event occurs. This latter tool may make use of a scripting language to define high-level actions that can be taken in response to an event, or it may offer a set of pre-defined actions that can be selected and arranged by end-users, or it may invoke named programs that can be implemented using a choice of programming languages. The exact nature of these actions is an open issue for this research project, however it is likely that the second option will be used as a starting point, so that non-technical stakeholders of the digital library can meaningfully participate in workflow creation and not depend on a programmer to script the actions that they require.

We will base the proposed workflow system on an infrastructure technology known as event notification systems. Event notification, in general, is based on the concepts of events, producers/consumers of events, and event subscriptions/notifications. Typically, an event is a set of attribute/value pairs. Producers publish events to an event server which routes these events to consumers based on their subscriptions. One benefit of this arrangement is that producers are completely unaware of the location of interested consumers (and are thus not dependent on these consumers in any way). Likewise consumers are unaware (and not dependent on) producers. This arrangement can lead to significant benefits. For instance, the C2 architectural style makes use of these characteristics to provide substrate independence in software architectures [23]. Other advantages include:

- Producers and consumers focus only on events meaningful to them. They have no need to understand the entire event space being managed by the event system. They are, thus, straightforward to create and configure.
- Event systems make efficient use of a network. When an event is produced, only those consumers who subscribed to the event are notified.
- If several event servers are used (which is, for instance, possible with the Siena system [11]), the routing of events can be further optimized (e.g. an event is only sent to an event server if it has clients that are interested in that event). This facilitates the use of an event notification system across a wide-area network.
- The publish/subscribe model enables dynamic service discovery. For instance, a consumer can publish an event requesting a specific service. If there is a producer that provides the service, it will notify the consumer, and the consumer can subscribe to it.

The use of event notification systems as a core technology allows us to create events with attributes associated with them. Thus, a "new submission" event would contain attributes that specified information such as a paper's title, authors, keywords, abstract, submission date, etc. Furthermore, workflows will be able to specify filters over various events, to ensure that only those stakeholders interested in an event will receive notification of it. Thus, as mentioned above, the workflow of an associate editor will begin with a new submission event that filters out any submission that does not match the associate editor's areas of expertise.

Event-based workflows directly address some of the concerns raised in the Introduction. First, event notification systems can easily be deployed across multiple organizations, and with the rise of systems like Siena [11], can easily route events across these organizations in a timely manner even across wide-area networks. Second, activities are specified in terms of domain-specific events which is a more intuitive way, we argue, to construct workflows, since it closely matches what people already do when coordinating their activities, e.g. "I'll wait until I receive that phone call

from Dan before proceeding." Third, event-based workflows are more flexible than "step-oriented" workflows since event-based workflows simply specify the order in which events must occur for a particular workflow to make progress, but they do not require that only one workflow be active at a time nor do they force a particular stakeholder to do activities in any particular order. One major problem with workflow technology is that some systems attempt to "enact" or "enforce" a particular workflow and some systems will go so far as to make it impossible for a user to work on task B unless they have first completed task A. Users rebel against such systems [22], and will take their work "off-line," outside of the system's control, to get work done. The proposed event-based system will allow users to work on tasks in any order, but may require that they "announce" the completion of task A (see below) before they "announce" the completion of task B. Finally, event-based workflows allow multiple organizations to participate in the same activity, support alternative organizations being responsible for the same step of an activity, and support different steps being associated with the same step in some larger workflow.

To illustrate, consider our example of reviewing new submissions to a digital library on literature that can receive submissions on either 18th century literature or 19th century literature. University A is responsible for reviewing the 18th century literature submission, while university B is responsible for the 19th century literature submissions. The journal has specified a very simple workflow for its review process:

```
NewSubmissionReceived [SubmissionAccepted | SubmissionRejected]
```

University A has specified the following review workflow:

```
NewSubmissionReceived, [FacultyReview1Received|FacultyReview2Received]+,
[SubmissionAccepted | SubmissionRejected]
```

University B has specified the following review workflow:

```
NewSubmissionReceived, GradStudentReviewReceived, [SubmissionAccepted |
SubmissionRejected]
```

Both universities participate in the digital libraries workflow since they "share" events. That is, the review workflows of the universities begin with the NewSubmissionReceived event which they share with the digital library's workflow. However, university A has specified interest only in NewSubmissionReceived events which have the keyword "18th Century Literature" (not shown), while university B has specified interest only in 19th century literature NewSubmissionReceived events. Thus, when a new submission is received by the digital library, the event filters will ensure that the correct university gets "activated" to review the submission. Once activated, the university uses its own workflow to review the submission, so University A will require that at least two faculty members review the submission, while University B will require that at least one graduate student review the submission. (Whether or not the universities should use the same reviewing standards is an issue that is out-of-scope for the workflow system. That issue must be resolved by the librarians and universities outside of the workflow system, however the workflow system may facilitate such a discussion since the differences in workflows will be made explicit by the workflow system.) Once the internal review cycle is complete, the university will generate an event that is once again shared by the digital library, either the submission is accepted or rejected. Once that event is generated, the focus of control for the workflow switches back to the digital library, who may have other workflows that activate to either inform the authors of the outcome of the review process and, if accepted, begins the process of incorporating the submission into the collection.

3.1.2 Architecture of Proposed Workflow Technology

The architecture of our event-based workflow system is shown below. It consists of an event notification layer, a workflow management layer, and a tool layer. The tool layer is split into design-time and run-time sections. Note: tools are accessed via a Web browser and the workflow management layer is managed by a set of services provided by a set of Web servers. Integrating these layers with elements of the Web architecture (e.g. Web servers and Web browsers) is critical to achieving our goals of producing lightweight, Web-based workflow technology. We provide more detail on each of these layers below.

Event Notification Layer. The event notification layer is responsible for delivering events between the organizations of the multiple stakeholders of a digital library, even if these stakeholders are separated by a wide-area network. The basic capabilities required by the proposed workflow technology is the ability to define events, support the publication (or generation) of events, support the ability for clients to register interest in events, and support the ability to route event notifications from publishers to all interested subscribers once an event occurs. In this sense, our workflow system should be independent of the particular event notification system used to implement the above architecture, since

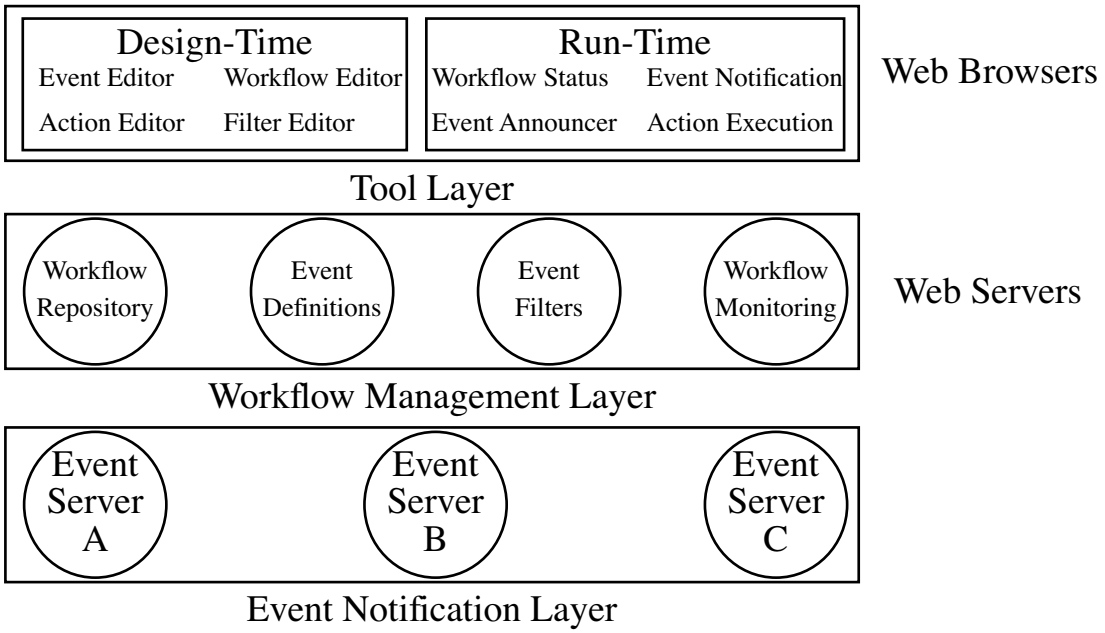

Figure 3-1. The architecture of the proposed workflow technology. The event notification layer is based on an Internet-Scale event notification system (Siena [11]) that ties multiple stakeholders together. A set of Web-based services provide a workflow management layer in which the persistence and linking of event-based workflows occurs. Finally, a tool layer, consisting of design-time and run-time tools provide stakeholders with the capability to construct, manage, and evolve event-based workflows.

most event notification systems support these capabilities. We have selected the Siena event notification system [11] to serve as an experimental platform for developing the proposed workflow technology for several important reasons. First, Siena is an Internet-scale event notification system, and thus has direct support for event notification across wide-area networks. Second, Siena's event servers and client APIs are available in a Java implementation, which lowers their entry barrier to use. This increases the chances that multiple stakeholders will be able to deploy this service in their computing environment. We intend to package Siena such that it is installed automatically when the proposed workflow technology is installed. This goal should make the existence of the event notification layer transparent to the users of the proposed workflow technology. Finally, reusing existing technology at this level will save at least a year's worth of work that would otherwise have to be performed and allows us to focus on the meat of the proposed research which is to develop workflow technology tailored to the digital library domain.

Workflow Management Layer. The workflow management layer is a server-side set of workflow services that enable the creation, manipulation, and persistent storage of events and event-based workflows. In addition, this layer dynamically manages the event filters for a particular stakeholder based on the current state of their activated workflows. That is, if a workflow consists of the simple pattern "A, B", then the workflow management layer will automatically register interest in events of type A when the workflow is activated. Once an event of type A is received, it will de-register interest in events of type A and register interest in events of type B. Finally, this layer is responsible for monitoring the status of each activated workflow. In particular, if an event for a particular workflow is received, the workflow management layer will update the activated workflow to record the current state, update the event filters to prepare for future events, and invoke any actions associated with the received event (which may include notifying the user that the event has occurred).

Current open issues for the workflow management layer include: 1) The nature and format of the event and workflow definitions, including the ability of workflows to reference and incorporate events. 2) The design and implementation alternatives of the proposed workflow services. 3) The ability to notify users of events, when those events are being received on the server but the users must be notified on the client.

We intend to explore the use of XML to encode workflow and event definitions. The primary reason for using XML is extensibility. While we intend to deliver a certain set of tools for the tool layer, the use of XML in the workflow man-

agement layer allows the tool layer to support an open set of tools. Existing and new third-party applications can be incorporated into the tool layer and applied to the event-based workflows stored in the workflow management layer as long as the tools support the XML-based event and workflow definition formats. A secondary reason for the use of XML is the wealth of tools that exist for editing and manipulating XML, including XSLT [12] that enables translating XML-based information into HTML for presentation by Web browsers. (XSLT is a suitable candidate for implementing a rapid prototype of the "Workflow Status" tool described below.)

With respect to services, the research will focus on defining a stable set of interfaces that can be used to access the services of the workflow management layer. It will then be possible to make use of any number of Web technologies such as Java servlets, CGIs, or even Microsoft "dot net" components [1] accessed via SOAP [2] to implement the services represented by those interfaces.

Finally, with respect to the third open issue, we will explore the design and implementation of "always active" Web clients that can be notified asynchronously when an event arrives. The primary goal of this aspect of the research will be achieving a low entry barrier to use with these clients. The event notification capability must be easy to activate, must stay out of the user's way when not in use, and must not hog an undue amount of computing resources. Otherwise, users will refuse to use them.

Tool Layer. The tool layer represents the user interface to the proposed workflow technology. It will consist of a set of design-time and run-time tools, accessed via a Web browser, that allow users to create, manipulate, and participate in digital library workflows. Design time services will consist of editors that allow users to create, manage, evolve, and share event and workflow definitions. The event editor will define the name and attributes of domain-specific events, and will feature the ability to group events into packages. The workflow editor will enable the construction of event-based workflows by providing a "drag-and-drop" environment for organizing events into patterns. The filter editor will allow users to specify specific conditions for when they are interested in particular events. Thus, in the example above, University A would use the filter editor to specify that they are only interested in NewSubmissionReceived events that contain a keyword attribute with the value "18th Century Literature." (Note: in an implementation, the filter editor may be directly incorporated into the workflow editor. At this stage of the research, we are identifying required functional capabilities, not specific tools.) Finally, an action editor will allow users to specify the actions that should occur when a particular event is received. With respect to run-time services, tools will be provided that record the status of a particular workflow (for instance, a visualization of the workflow that highlights the last event received), notify users of events (discussed above), allow users to announce events, and execute the actions associated with a particular event.

While visualization of a workflow should be straightforward, the last two services require special attention. The event announcer is a tool that allows a stakeholder to generate an event notification. The existence of this tool is required, since it is impossible to automate the generation of all workflow events. For instance, it may be possible to automate the generation of a NewSubmissionReceived event, since a digital library may have a Web-based submission system. Once a user clicks on a "Submit Submission" button of a Web-based form, a CGI script (or similar) can access an interface of the workflow management layer to generate a NewSubmissionReceived event with its attributes set to values contained in the Web-based form. However, it may be impossible to automate the GradStudentReviewReceived event, since the graduate student may send the review to an editor via e-mail. In this case, either the grad student or the editor needs to "announce" that the GradStudentReviewReceived event has occurred. The event announcer tool will provide an interface to facilitate this task, and can provide features such as automatically selecting the event that needs to be announced based on its knowledge of the activated workflows and providing forms to fill out the attributes associated with a particular event. Once the event has been filled out by the user, the user can click a button and the event announcer will publish the event and allow the associated workflow to proceed.

With respect to the tool that executes actions in response to an event, the requirement that this tool be "Web-based" may have to be relaxed. The reason for this is that it may not be possible to invoke some desirable actions from within the "sandbox" environment that most Web-based services execute. In other words, Web-based services, especially those that are delivered in the form of Java Applets, are not allowed to access a user's file system or invoke applications that exist outside the Web browser. Since we envision this tool being able to perform actions like sending e-mail and transferring files, sandbox restrictions would be too limiting. We therefore intend to strike a balance between the capabilities of this tool and the security concerns that the sandbox was meant to address. In particular, we intend to define a small set of primitive capabilities that can be activated and deactivated by individual users and an extension mechanism that will allow them to add to the set of actions that can be performed by the tool. The exact nature of this

extension mechanism is an open issue for the proposed research, but desirable characteristics include the ability to invoke tools in a platform independent way (which may require the use of a tool server [13]), the ability to incorporate information from events as parameters to these tools, and the ability to incorporate newly defined actions into the design-time action editor.

3.1.3 Evaluation Plans

We intend to evaluate the utility of the proposed workflow technology by designing, building, and deploying prototype tools in our three evaluation environments, collecting feedback from real users, and evolving our designs, functional capabilities, and tools in response. We intend to apply the same basic methodology to each of our three evaluation environments. During the initial phase of the research, we shall have our evaluation partners (the Co-PI at USRA, the Senior Person at NCAR, and Dr. Sumner at JIME) collect metrics and data on how current activities are performed without the presence of workflow technology. As basic workflow technology is delivered by our initial prototypes, we will measure the impact that it has on the previously measured activities. We hypothesize that the prototype will have an initial negative effect on efficiency while the stakeholders learn to use the new tools and the inevitable bugs are discovered and fixed in the prototype software. As the research continues, however, and more sophisticated functionality is delivered to the stakeholders in an incremental fashion, we expect to see the efficiency of the activities to increase and the overall time that is needed to complete workflows to decrease. In addition, we expect that the stakeholders will demonstrate an increased awareness of the global state of the workflows that they participate in. We expect that such awareness will manifest itself in the ability of the stakeholders to identify problems or bottlenecks with their workflows and to use this information to evolve their workflows to correct those problems.

3.1.4 Deliverables

The deliverables of the proposed research include the following:

1. Papers that document the information identified by the three goals listed at the end of the Introduction. These papers, which will include information on the developed workflow technology and the results of our evaluations, will be submitted for conference and journal publication.
2. Software that implements the functional capabilities identified for the workflow management and tool layers. This software will include the XML Schemas that define the XML-based data formats of the workflow management layer, the Web-based services that implement the repository and monitoring functions of the workflow management layer, and the applications that implement the functional capabilities of the tool layer.
3. Course materials developed for software engineering courses that present the results of the research as an addition to the existing material on workflow technology. In particular, the new material will present insights into how workflow is different in the digital library domain, and how workflow techniques and technology must be adapted to successfully support digital library workflows.

3.2 Research Plan

This project will be carried out over two years with assistance from two graduate research assistants (GRAs). The GRAs will work on each of the major research tasks defined above. The PI will manage the work of the GRAs and facilitate the gathering of requirements from the Co-PI and Senior Person. Note, the GRAs will interact with these people as well, however their interactions will focus more on clarification of information gained from the PIs requirements gathering activities. In addition, the PIs will ensure that the students are co-located and that regular project meetings are held to keep the project on track. The following plan describes the work to be carried out in each year, along with work allocations. Student work allocations assume management by the PI.

The tasks are roughly divided such that design-time capabilities are designed and developed in year 1, while run-time capabilities are designed and developed in year 2.

3.2.1 Year 1

- Collect additional requirements for the proposed workflow technology (PI, Co-PI, Senior Person)
- Collect metrics on existing activities within the evaluation environments (Co-PI, Senior Person, Dr. Sumner)
- Design XML-based data formats for events and workflows (GRA 1)
- Design and rapid prototype repositories for these formats and their associated interfaces (GRA 1)
- Design and rapid prototype design-time tools including the event, workflow, filter, and action editors (GRA 2)
- Deploy repositories and tools into evaluation environments for feedback (PI, Co-PI, Senior Person, Dr. Sumner)
- Use feedback to evolve design-time interfaces, data formats, repositories, and tools (GRAs)

3.2.2 Year 2

- Collect metrics on the ability of stakeholders to use the design-time tools (PI, Co-PI, Senior Person, Dr. Sumner)
- Design and rapid prototype workflow monitoring service of the workflow management layer, including the ability to activate workflows, monitor workflow status in response to incoming events, and update event filters automatically (GRA 1)
- Design and rapid prototype run-time tools including the workflow status monitor, the event announcer, the action execution tool, and the event notification tool (GRA 2)
- Deploy run-time services and tools into evaluation environments for feedback (PI, Co-PI, Senior Person, Dr. Sumner)
- Collect metrics on the impact of workflow technology on the activities of the evaluation environments (PI, Co-PI, Senior Person, Dr. Sumner)
- Use feedback to evolve run-time services and tools (GRAs)
- Develop new course materials on supporting workflow in digital library environments (PI)

4 Background and Related Work

In this section, we briefly examine background information and related work in the areas of event notification systems and workflow management systems.

4.1 Event Notification Systems

The proposed workflow technology builds on top of the services provided by event notification systems. In fact, as explained in Section 3.1, the proposed workflow technology is independent of any particular event notification system, such that our prototypes can make use of any event notification system that meets the specified requirements. Event notification systems were first employed to support tool integration in software development environments. One of the first systems to employ this approach in a local-area network setting was Field [20]. Tool integration via events was also a part of Hewlett Packard's SoftBench environment [10]. In recent years, event notification systems have been extended to explore issues related to events across wide-area networks [11] and enabling project awareness [14, 15, 19]. For the proposed research, we intend to make use of the Siena event notification system [11] due to its ability to support Internet-scale event notification services (which is ideal for coordinating the actions of the stakeholders of a digital library) and the fact that both its client and server packages have Java implementations, which support our goal of producing lightweight workflow technology.

4.2 Workflow Management Systems

Workflow Management Systems are tools for storage and manipulation of process models that facilitate the completion of some task, such as the development of a software system or the processing of a travel authorization form. These process models act as artifacts for the coordinated activity. Such artifacts can represent, to a limited degree, the procedures and rules that mediate an effort, as well as some aspects of the division of work. This representation is typically limited to the most abstract and restricted sequence of independent work steps. Workflow systems typically do not venture into more complex forms of interaction. However, the results of these more complex interactions can be incorporated to a certain extent via the transformations of artifacts performed by users as a result of decisions made outside the scope of a workflow system proper, e.g., as a result of conversations held over lunch [6]. There are three interaction paradigms used in workflow systems:

- **Task-oriented:** This interaction style involves the use of agendas. Agendas manage lists of relevant tasks for each user, as e.g. in SPADE-1 [5].
- **Document-oriented:** Interaction is achieved in these systems via documents and document services. In Merlin [16], for instance, a work context graphically displays the relevant documents available to and associated with each user role.
- **Goal-oriented:** Interaction in this paradigm is centered around a list of goals to be accomplished. In Marvel, for instance, these goals represent currently active rules that can be applied to the state of the process [17].

In each of these paradigms, events play a secondary role. Thus, a novel aspect of this research is to examine the capabilities enabled by making events first-class citizens in the proposed workflow technology. Note, while various workflow systems are Web-enabled—for instance, Endeavors [8] and OzWeb [18]—few, if any, have as a goal to make the entry barrier to use of the workflow system as low as possible. Thus, another novel aspect of the proposed research is

to discover the right balance of techniques and technologies to create a truly lightweight workflow system, that can be deployed across a wide spectrum of digital library stakeholders.

5 Potential Impact on Higher Education

The proposed research has the potential to impact higher education at the undergraduate and graduate levels in the domains of computer science and Earth system science.

With respect to computer science education, there are several significant opportunities. First, software engineering graduate students will have the opportunity to work and interact with the stakeholders of the NCAR Web portal which is situated in a large-scale, complex, "real-world" computing environment. (NCAR is a fifteen minute drive from CU.) This setting will directly expose them to issues of scale and heterogeneity and will require them to work collaboratively with NCAR personnel to both understand and augment the complex workflows that surround the Web portal. In addition, they will be exposed to the nature of multidisciplinary research, since they must gather requirements from and build tools for researchers who are not computer scientists. The graduate students will also benefit from the feedback that they receive from interacting with the Co-PI at USRA with respect to supporting the complex workflows of the JESSE electronic journal. This interaction will occur primarily through e-mail and telephone.

Second, the proposed research will impact the graduate-level Foundations of Software Engineering course (CSCI 5828) and the undergraduate-level Software Methods and Tools course (CSCI 3308) at CU by providing insight into the unique nature of workflow in a digital library and how traditional workflow technology must be modified to adapt to these conditions. In addition, it can lead to new instruction, course work, and class projects in which students learn about and apply cutting edge technologies such as XML and event messaging systems. More importantly it can demonstrate how these enabling technologies can influence the analysis and design process since the proposed workflow technology is being designed specifically to exploit the existence of these technologies rather than attempting to build everything from scratch. Note, while the proposed research will impact the courses of the University of Colorado first, the PI places all of his developed course materials on the World Wide Web (<http://www.cs.colorado.edu/users/kena/classes/>), where they are accessible to educators from other universities.

With respect to undergraduate education in Earth system science, JESSE's main goal is to impact Earth system education by providing ready access to quality learning materials for educators and students at all levels, beginning with the undergraduate community. JESSE is providing a model early on for interdisciplinary electronic peer review and publication of learning resources from a broad range of disciplinary scientists engaged in the collaborative development of Earth system and global change science. Flexible, lightweight workflow enhancements will improve this process, and make it easier for authors to share their research and education resources.

Graduate teaching assistants often author and implement learning resources which are used to supplement undergraduate courses. JESSE encourages students to contribute and will provide special recognition and annotation of student submissions. Workflow enhancements developed here will be especially important to foster a meaningful level of student involvement in publishing their work. Subject to the same review process, publication of these student-authored resources will offer recognition that will benefit the professional careers of the authors.

Of special interest will be the development of workflow enhancements that encourage and assist resource submissions from minority serving institutions and under represented groups. JESSE is making a special effort to reach 2 year institutions, recognizing their role in educating large numbers of students who go on to four year college programs. JESSE will be a means by which authors at these institutions can share the teaching materials that they have developed and receive peer and academic recognition for their efforts. Since JESSE fills a specialty niche in the professional publication of Earth system education content, a new market of educational contributors will be tapped, and many authors may find it easier and more relevant to publish educational contributions in JESSE than in the more traditional, research oriented journals. Lightweight web-based workflow tools will be an important part of making this possible.

The success of the NSDL as a superstructure for digital libraries of specific scientific disciplines requires collaboration across disciplines in order to develop an orderly array of resources that ensure access, utilization, communication and flow of information in serving both the disciplinary and broader interests of the nation. The workflow tools developed here will enhance the process to the benefit of the authors and readers. The interdisciplinary activities of JESSE and the experience gained in developing Earth system education resources and more efficient workflow processes will help develop pathways and bridges of collaboration and utilization among higher education institutions, a service ultimately needed for fulfilling the functions envisaged for NSDL.

6 Results from Prior NSF Research

6.1 Kenneth M. Anderson

The PI from the University of Colorado, Boulder started his first NSF-funded research project (CCR-9988517: "Supporting Information Integration in Large-Scale Software Development") in September 2000. While some progress has been made on this project, the results are preliminary and do not directly impact the research of this proposal.

6.2 Martin Ruzek

6.2.1 EAR-9907764

"A Peer Reviewed Electronic Journal of Earth System Science Education Resources (JESSE)" $49,967, Period of Performance - 9/1/99 - 12/31/00

This project defined the structure and process of JESSE as a prototype effort, leading to support for adequate staffing to implement the Journal on an operational basis. Results include:

- Extended awareness of the JESSE concept and sought community participation through concept papers and abstracts presented at national and international meetings (see below), as well as informal presentations to the DLESE Steering Committee, the NASA Langley DAAC Users Working Group, etc.

Ruzek, M., D. R Johnson, M. Kalb "Peer Review for Digital Earth: A Journal of Earth System Science Education" in Towards Digital Earth - Proceedings of the International Symposium for Digital Earth, Beijing, China November 29 - December 2, 1999 pp. 235 - 239

Johnson, D.R., Ruzek, M., Kalb, M., "A Vision for the Journal of Earth System Science Education (JESSE) and the Collaborative Development of Educational Resources" Eos Trans. AGU, 80(46) Fall Meet. Suppl., F119, 1999

Ruzek, M., Johnson, D.R., Kalb, M., "What Will Change the Nature of Geoscience Education and Shape its Future in the Coming Century?" Eos Trans. AGU, 80(46) Fall Meet. Suppl., F116, 1999

Johnson, D.R., Ruzek, M., Kalb, M., "Earth System Science Education in the New Millennium: Learning Modules and Peer Review", American Meteorological Society 9th Symposium on Education, pp 97-100, Long Beach, CA January 9-14, 2000 Johnson, D., R. Ford, M. Ruzek, M. Kalb "JESSE: Peer Review for Digital Library Content" Eos Trans. AGU, 81(19) Spring Meet. Suppl., S67, 2000

Ruzek, M., D. R. Johnson, M. Kalb "An Electronic Peer-reviewed Journal of Earth System Science Education" abstract and paper for IGARSS 2000, Honolulu, HI

- Initiated dialog with professional societies and organizations (AMS, AGU, CORE, IRIS) seeking areas of common interest and mutual benefit, first recognizing JESSE's unique contribution as a valid outlet for member publication, then assisting JESSE with links and web referrals for dissemination of JESSE content and solicitation of contributions from their membership.
- Initiated the design of the organizational database to manage the flow of learning resources through the review and publication process. Database structure is informed by the IMS metadata standards.
- Refined the JESSE process and structure in coordination with the needs and plans for DLESE, especially adopting the review criteria as recommended by the DLESE Collections Committee.

6.2.2 DUE-0085837

"Implementing an Electronic Peer-reviewed Journal of Earth System Science Education Resources (JESSE): A Pathfinder for SMETE Resource Peer Review)" $249,861, Period of Performance - 10/1/00 - 9/30/01

This ongoing project is implementing JESSE with plans for release of the first issue in Summer 2000.
See <http://jesse.usra.edu/>.

Progress to date includes:

- Convened the JESSE Editorial and Advisory Board in Nov. 2000 for input and advice.
  See <http://jesse.usra.edu/testing/pemeetingnotesnov2.html>.
- Met with Gerry Hanley of the MERLOT project in December, 2000 to discuss ways to work together towards effective peer review for NSDL

- Participated at DLESE working group meetings representing JESSE
- Prepared detailed JESSE guidelines for contributors and guidelines for reviewers
- Installed the Digital Document Discourse Environment (D3E) on the JESSE server customized to meet JESSE's needs.
- Developed a partnership with John Weatherley of the University of Colorado to help refine the D3E review process, working with him to assess the reviewer/author experience
- Revised document flow within JESSE and identified the need for a redesigned interface for peer commentary
- Initiated the testing of the end to end review process with a real submission in March, 2001, and identified procedural changes to aid reviewers
- Continued development and testing of the web-based database to organize author, reviewer and resource information
- Developed and implemented the prototype JESSE web site that the reader will use, and continue to work on web tools for managing the submission and review process, as well as community commentary.
- Ruzek served as convener of a special session on Peer Review and JESSE at the DLESE Leadership Workshop, Bozeman, MT July, 2000

Johnson, D.R., Ruzek, M., Ford, R., "Implementing an Electronic Peer Reviewed Journal of Earth System Science Education Resources (JESSE): Pathfinder for SMETE Resource Peer Review" Eos Trans. AGU, 81(48) Fall Meet. Suppl., F295, 2000

Ford, R., M.Ruzek, D, Johnson "The Journal of Earth System Science Education" oral presentation and poster, AAG annual meeting, February, 2001, New York

Johnson, D., M. Ruzek, J. Weatherley, "The Journal of Earth System Science Education: Peer Review for Digital Earth and Digital Library Content" abstract submitted for education session at Spring AGU, Boston, May, 2001

6.3 Aaron Anderson

The Senior person from NCAR has never been a principal investigator on an NSF-funded research project.

7 Summary of Proposed Research and Deliverables

The proposed research has identified a need for new techniques and tools to support digital library workflows. In particular, certain characteristics of digital libraries, such as the diverse and distributed nature of digital library stakeholders, make applying traditional workflow technology difficult, if not impossible. Based on a set of example workflows, requirements were identified for a new type of workflow technology that is lightweight (easy to adopt), flexible (easy to adapt) and Web-based (easy to deploy). A research plan was presented to design and develop the proposed workflow technology and the research issues that must be confronted in performing this work were identified and discussed. The proposed research intends to provide the following deliverables:

1. An understanding of the characteristics of digital library workflows
2. An understanding of the types of services needed to support and coordinate these workflows.
3. An understanding of the heterogeneity encountered in digital libraries and the techniques required to create workflow technology that can effectively handle this heterogeneity and provide a low entry barrier to use.
4. An understanding of how improved support for workflow can impact the ability of a digital library to meet the needs of its users.
5. A prototype of the proposed workflow technology, including implementations of the event notification, workflow management, and tool layers.
6. Results of the effectiveness of the developed workflow technology as measured by deploying the prototypes into the "real-world" environments of three digital libraries (or, rather, environments that share many of the same characteristics of digital libraries.)
7. New workflow-related materials for software engineering classes at the graduate and undergraduate levels.

D. References Cited

1. Microsoft .NET Home Page. <http://www.microsoft.com/net/default.asp>.
2. Simple Object Access Protocol. <http://www.w3.org/TR/SOAP/>.
3. Ambriola, V., Conradi, R., and Fuggetta, A. (1997). Assessing Process-Centered Software Engineering Environments. *ACM Transactions on Software Engineering and Methodology,* 6(3)**:** 283-328.
4. Anderson, K. M. (1999). Issues of Data Scalability in Open Hypermedia Systems. *The New Review of Hypermedia and Multimedia,* 5**:** 151-178.
5. Bandinelli, S., Di Nitto, E., and Fuggetta, A. (1996). Supporting Cooperation in the SPADE-1 Environment. *IEEE Transactions on Software Engineering,* 22(12).
6. Barthelmess, P., and Anderson, K. M. (2001). A View of Software Development Environments Based on Activity Theory. *Computer-Supported Cooperative Work,* to appear in the Special Issue on Activity Theory.
7. Berners-Lee, T. (1996). WWW: Past, Present, and Future. *Computer,* 29(10)**:** 69-77.
8. Bolcer, G. A., and Taylor, R. N. (1996). Endeavors: A Process System Integration Infrastructure. In *Proceedings of the 4th International Software Process Workshop*, pp. 76-85. Brighton, U.K. December 1996.
9. Bolcer, G. A., and Taylor, R. N. (1998). Advanced Workflow Management Technologies. *Software Process: Improvement and Practice,* 4(3)**:** 125-171.
10. Cagan, M. R. (1990). The HP SoftBench Environment: An Architecture for a New Generation of Software Tools. *Hewlett-Packard Journal,* 41(3)**:** 36-47.
11. Carzaniga, A., Rosenblum, D. S., and Wolf, A. L. (2000). Achieving Scalability and Expressiveness in an Internet-Scale Event Notification Service. In *Proceedings of the 19th ACM Symposium on Principles of Distributed Computing*.
12. Clark, J. (1999). XSL Transformations (XSLT) Version 1.0 W3C Recommendation, 16 November 1999. <http://www.w3.org/TR/xslt.html>.
13. Dossick, S. (1998). The Rivendell Tool Server. <http://www.psl.cs.columbia.edu/Rivendell/>.
14. Fitzpatrick, G., Mansfield, T., Kaplan, S., Arnold, D., Phelps, T., and Segall, B. (1999). Augmenting the Workaday World with Elvin. In *Proceedings of the 6th European Conference on CSCW*, pp. 431-450.
15. Fuchs, L. (1999). AREA: A Cross-Application Notification Service for Groupware. In *Proceedings of the 6th European Conference on Computer Supported Cooperative Work*, pp. 61-80.
16. Junkerman, G., Peuschel, B., Schäfer, and Wolf, S. (1994). MERLIN: Supporting Cooperation in Software Development through a Knowlege-Based Environment. Pages 103-130 in A. Finkelstein, J. Kramer, and B. Nuseibeh, Eds., *Software Process Modelling and Technology*. Research Studies Press Ltd.
17. Kaiser, G. E., Barghouti, N. S., and Sokolsky, M. H. (1990). Preliminary Experience with Process Modeling in the Marvel Software Development Environment Kernel. In *Proceedings of the 23rd Annual Hawaii International Conference on System Sciences*, Vol. II — Software Track, pp. 131-140. Hawaii, USA.
18. Kaiser, G. E., Dossick, S., Jiang, W., and Yang, J. J. (1997). An Architecture for WWW-Based Hypercode Environments. In *Proceedings of the 19th International Conference on Software Engineering*, pp. 3-13.
19. Prinz, W. (1999). NESSIE: An Awareness Environment for Cooperative Settings. In *Proceedings of the 6th European Conference on Computer Supported Cooperative Work*, pp. 391-410.
20. Reiss, S. P. (1990). Connecting Tools Using Message Passing in the Field Environment. *IEEE Software,* 7(4)**:** 57-66.
21. Ryder, M. (1996). Print vs. Online Scholarly Publishing: Notes and Reflections on the Peer Review Process. *International Journal of Media and Communication Studies,* Editorial Board Comments.
22. Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Computer Communication*. Cambridge University Press, New York.
23. Taylor, R. N., Medvidovic, N., Anderson, K. M., Whitehead, E. J., Jr., Robbins, J. E., Nies, K. A., Oreizy, P., and Dubrow, D. L. (1996). A Component- and Message-Based Architectural Style for GUI Software. *IEEE Transactions on Software Engineering,* 22(6)**:** 390-406.