Workflow and Electronic Records Capture

Rosemary Pleva Flynn
Electronic Records Project Archivist
Indiana University Archives
Last revised May 8, 2002

Introduction
In recent years, many of the information technology (IT) trade publications have been reporting on the automation of business processes using workflow engines. When combined with a web portal for access, a workflow engine can help provide information and services both inside and outside an organization. How can archivists and records managers leverage these new technologies to capture, maintain, and preserve electronic records? As Chen notes in “The Paradox of Digital Preservation,” an article in Computer, a journal of the IEEE Computer Society, “In organizations, workflow processes continuously create documents and records. Starting from creation and ingestion, we should integrate the workflow process with the preservation process: appraisal, verification, maintenance and, eventually, retirement.”

At Indiana University, as part of the second phase of the IU Electronic Records Project, the project team has been exploring the possibility of using a workflow engine to capture records as they are being created in the electronic environment and then transferring them to an electronic recordkeeping system for their maintenance and preservation. The team has been involved with the design and implementation of EDEN (Enterprise Development Environment), the shared infrastructure that provides services like a workflow engine to university applications, and OneStart, IU’s campus portal that is user interface.

What Is Workflow?
Workflow is "the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant [human or machine] to another for action, according to a set of procedural rules." In other words, it is a set of logical steps known as activities that are either executed by a human user or by a machine.

A workflow engine is the software that facilitates the automation of the business process. In the early days of workflow, the engine was concerned with the electronic delivery of work to people. Now, the business process itself can be fully automated. The workflow engine can perform activities such as updating other systems automatically, triggering application to produce other documents, even making decisions based on programmed criteria. This automation is particularly useful with repetitive activities, which few people enjoy. While complete automation may not be practical for all business processes, the

---

1 For more information on workflow technology, see the Workflow Management Coalition's website at http://www wfmc org/. If you would like more information on portals, particularly portals in higher education, visit the research report on portals from Eastern Michigan University's University Computing at http://ucinfo emich edu/aboutUC research portals cfm. It includes an extensive list of references.
3 The original published name for the portal was MyIU. As stated in the FAQ on the project’s website, “The MyIU name reflected the current convention in portal naming. While it's easy to guess, it also left out our colleagues who are Purdue faculty and students.” Online at http://www indiana edu/~onestart project/index htm
4 From http://www e workflow org/, a workflow portal sponsored jointly by the Workflow Management Coalition (WfMC) and the Workflow And Reengineering International Association (WARIA)
goal of workflow is Straight-Through-Processing where human contributions are kept to a minimum and are generally related to management of exceptions within the process. Effectively implemented workflow offers a variety of benefits to an organization including reduced operating costs, improved productivity and services, and better conditions for employees.

**Workflow at Indiana University**

In 1998, Indiana University released its five-year IT Strategic Plan. Among the recommendations and proposed actions were the replacement of several enterprise wide applications and the re-engineering of other legacy applications to web-enable them. Also, like many other colleges and universities, IU is implementing PeopleSoft’s Human Resource Management System (HRMS) and Student Information System (SIS). The implementation of these software applications provided an excellent opportunity to integrate all of the enterprise applications at IU. The integration plan involved a transaction-processing environment that would provide access to these applications through a coordinated, unified front end and an infrastructure made up of components that would be shared among applications including a workflow engine and a global inbox for administrative messages.

The requirements gathering began in March 2000 with a Joint Application Design (JAD) session involving IT developers and managers and several representatives from key functional areas including Human Resources, Timekeeping, Student Information, Financial Management, Accounts Payable, Purchasing, Electronic Research Administration, and the Indiana University Information Environment. From this session and additional meetings with these functional areas during the summer of 2000 a vision for the portal emerged. It would provide a unified front end to IU services with single sign-on and authentication and 24x7 availability, role-based customization, usability-tested personalization features, application integration, an adaptive user interface, and a completely user-centered environment. It would also make use of a component-based design (CBD)/Web services methodology and standards approach. This approach allows for the development of a shared infrastructure.

OneStart can be referred to as a “next generation” portal since it is more than an information portal. Not only is it flexible and responsive to change, it utilizes a distributed model for content creation meaning OneStart provides the framework for others to publish their services and content. The capturing of records in such a distributed arena can be a nightmare for records managers and archivists. This is where the automatic capture of electronic records using the workflow engine will be useful.

As mentioned previously, OneStart and EDEN are being developed using component-based design. A component is a specific piece of enterprise functionality that can be reused in future development and integration. As shown in Figure 1 below, some of the components that were originally part of EDEN included a workflow engine, recordkeeping, security, and an inbox. A key difference between components and objects is that components clearly separate specification from implementation allowing for easier reuse. Components also have published interfaces that should be independent of their implementation in order to facilitate integration.

---


7 This overview of component-based design (CBD) comes from a presentation given at CUMREC 2001, A Higher Education Technology Conference (an EDUCAUSE affiliate) by James Thomas and John Walsh, “Indiana University Has Embarked on a Journey to Create the "Next Generation" Web Portal,” which can be found at http://www.educause.edu/asp/doclib/abstract.asp?ID=CMR0105
A component-based approach has several benefits. First, it creates a repository of reusable business functions that also allows for the replacement of specific functions without affecting the rest of an application. Also, it aids rapid application development by assembling existing components and services. CBD can improve the agility, flexibility, and scalability of an application. Finally, as long as components agree upon the protocol to be used (the language and the subject), an application does not have to reach into another application's database for information. They only interact through their published interfaces.

The Conceptual Design for the IU Workflow Engine
The idea for the system-wide workflow engine was based on an engine used for routing documents in the current financial management system. For many applications, the workflow engine will be used to route e-docs. When the IT community and functional areas at IU talk about e-docs, they are generally referring to electronic forms that are connected to database records. These electronic forms are much like the traditional paper forms they will replace. E-docs usually do not include other file types such as scanned images although the engine will be able to accommodate these. The workflow engine will route e-docs for activities such as completion, approval, or notification.

The diagrams below illustrate a high-level conceptual design for the workflow engine and how it interacts with components like the portal, inbox, and other applications. Figure 2 offers an overview of the engine and other components including its interaction with a recordkeeping system. Figures 3-1 through 3-9 provide an example of how the components work together through one business sub-process. Since this is conceptual, very little concerning implementation will be discussed. Also, as far as EDEN is concerned, OneStart is just another application like the Financial Information System (FIS) or the Human Resources Management System (HRMS). However, it does have a special relationship since the portal is the user interface for the inbox and other EDEN services. For this reason, OneStart has been set to the left of EDEN in the diagrams rather than grouping it with the other applications.
The best way to explain how the engine and other components work together is to use an example involving a document from the HRMS. The HRMS will have a component that represents a particular document type such as a Personnel Action Form (PAF). The system will register routing rules with the workflow engine on how it should route. The workflow engine routes an XML representation of the document from node to node through the business process. The HRMS owns the data and is responsible for its storage and management. However, the workflow engine does tell the HRMS how that should be done in order to maintain the document's route. All communication between the HRMS and the workflow engine is conducted between their respective interfaces. The example has been broken down into a series of figures and descriptions.

The process begins when someone initiates a PAF for a new employee named Joe in the HRMS. Whether Joe's PAF needs to be routed to another node for completion or is ready to begin the approval process, the HRMS sends a representation of the document to the workflow engine when it is ready to be routed. (Figure 3-1)

The workflow engine initiates a discovery process to decide the node where the PAF needs to go based on the routing rules registered with the engine. Nodes will have an associated inbox, which may be for a person, machine, or another process. In this example, notice of the document's availability is sent to an employee's inbox by the workflow engine. (Figure 3-2)
The employee views the inbox with its the list of documents through the portal. (Figure 3-3)

Within the portal, the employee selects Joe's PAF from his or her inbox. (Figure 3-4)

The inbox contacts the workflow engine to get the document. (Figure 3-5)

The workflow engine will call on a HRMS component for a representation of the document. (Figure 3-6)
The HRMS component will request the workflow engine to check the status of the PAF (the document's state and what actions can be taken). (Figure 3-7)

The workflow engine will then provide a representation of Joe's PAF that is displayed through a OneStart channel. (Figure 3-8)

Once an action is taken, the document returns to the workflow engine, the action is recorded, and the process is repeated until the PAF reaches its final node. (Figure 3-9)

**Record Capture**

Expanding this conceptual design to include record capture capability involves the inclusion of a recordkeeping node in the business process. Routing rules would direct the document to that node when the document becomes a record of a transaction. As shown in Figure 4, the inbox attached to the node would have a conduit that passes the document into a recordkeeping system or repository where it can then be managed. Each document in the recordkeeping system would include the metadata attached at its creation in addition to all metadata gathered whenever it is routed through the engine. Since the record has been captured as part of its associated business process, archivists and records managers can be assured that the metadata also contains all of the appropriate contextual information that may be missing if captured at a later date.
IU does not have at this time, nor do they plan to have in the near future, a document management or recordkeeping system. Having identified a means for capturing records from the various university systems, the electronic records project team has explored a partnership with the Indiana University Information Environment (IUIE), IU's data warehouse, in order to provide a repository for these records. This would allow the team to make use of existing university resources until a more complete system can be put in place.

Conclusion
Using a workflow engine to capture electronic records and transfer them to a recordkeeping system is not a simple thing to do at this point. For example, time-consuming business process modeling must be done to determine when records must be captured. Also, extensive work must be done with individual applications to ensure that the appropriate metadata information is part of each document type. Once records are captured and transferred to a recordkeeping system, measures must be in place to ensure future access by application users who use these records sometimes on a daily basis. At a minimum, this involves developing additional interfaces between the recordkeeping system and the workflow engine in order to recall the records. In addition, duplicate records in the original applications still must be managed until they are destroyed. There are likely other issues and concerns that are not addressed in this brief conclusion. However, using a workflow engine to capture records is definitely worth pursuing as part of an overall electronic records program: few other solutions have been offered up to this point and the capture of records as they are created in the business processes does ensure the most complete records possible in terms of capturing content, context and structure.

From:  http://www.indiana.edu/~libarch/ER/
Indiana University Archives – electronic Recordkeeping – Project Whitepaper