Process-driven Management Information Systems - Combining Data Warehouses and Workflow Technology

Michael zur Muehlen
Department of Information Systems
University of Muenster
ismizu@wi.uni-muenster.de

Abstract

The use of workflow technology promises efficiency gains through the automation of manual routing, coordination and work distribution tasks. During the execution of workflows, state-changes of the workflow engine are recorded in a log file or database, the so-called audit trail. Combined with business object data, the audit trail provides exact and timely information about the operational behavior of an enterprise. In this paper we discuss the design on data warehouse applications that take advantage of workflow technology as an information source. We outline evaluation opportunities generated by the use of audit trail data and point out potential pitfalls with regard to data consistency and integrity.

1 Introduction

The use of workflow management systems improves the efficiency of business processes through the automated coordination of the data and resources needed for the execution of the single activities. Workflow management systems rely on a formal representation of the process logic that is designed as a workflow model during the development phase of a workflow application (also called *build-time*). During the execution phase of the workflow application (also called *run-time*), the workflow engine derives workflow instances from the generic workflow model and notifies workflow participants about pending activities through their work-lists (see e. g. [14]).

Workflow management systems have found widespread acceptance since the advent of this technology in the late 1980s. The Association for Information and Image Management (AIIM) estimates the worldwide revenue for workflow technologies to grow from $4.3bn in 2000 to $8.3bn in 2003 at a compound annual growth rate of 31% [6]. Especially in conjunction with document management technology, workflow systems are perceived as the enablers of office productivity gains through the elimination of manual routing and work distribution tasks [6]. Recently, workflow management systems have spread beyond the administrative environment and can also be found as embedded software components, that enhance existing application packages (e. g. ERP systems) as well as infrastructure components (such as application servers) with process management functionality [31].

Although workflow management systems can increase the process efficiency of an enterprise by as much as 150% [17], they do not necessarily lead to a more flexible organization. Since the introduction and deployment of a workflow-based information system architecture is very often a complex and time-consuming endeavor, it can be observed, that once this kind of architecture has been successfully deployed, many companies resist the urge to apply changes to the new system (an effect that can be observed at companies introducing ERP packages as well).

However, the complexity of workflow projects is only one cause for the reluctant attitude observed towards change management. An even more severe cause is the non-transparency of those relationships, that describe the effects of workflow changes on the organizational, technical, or process level. This missing transparency can be attributed to the lack of an integrated infrastructure for the gathering and presentation of performance indicators, that describe the behavior of a workflow-enabled organization and give advice on which parameters to adjust in order to increase the organizational efficiency.

During the run-time phase of a workflow application, the workflow engine generates information about the different state-changes on the process and activity level that is recorded in the so-called *audit trail*, which is either file-based or a database structure (cf e. g. [19]). Originally designed as a technical protocol for debugging purposes, the audit trail provides information about the execution of business processes at the operational level.

In theory, the use of audit trail information could enhance traditional enterprise controlling and management information systems by presenting key indicators about process performance as well as drill-down capabilities from single process instances to the business data connected with these instances. In practice, the reporting components of many workflow management systems offer only limited evaluation functionality. Especially the combination of workflow audit trail data with data warehousing technology is not supported by most workflow vendors, but left to extensions by the user.

Published in: Gavish, B. (Ed.): Proceedings of the Fourth International Conference on Electronic Commerce Research (ICECR-4), Dallas, TX, November 8-11, 2001., pp. 550-566

This paper aims at providing a framework for the design of data warehouse applications that are specifically targeted at the integration of workflow audit trail data. In the following section we develop a taxonomy of different analysis purposes for workflow audit trail data. We distinguish between workflow monitoring and controlling, and show, how both concepts relate to the workflow life cycle. In section 3 we focus on the design issues for a process information system based on workflow audit trail data. Following a phases model that spans data extraction, transformation, evaluation and presentation, we point out benefits and potential problems associated with the use of audit trail data for data warehouse applications. In section 4 we show the benefits of workflow-driven process information systems using a case study from an insurance company. A discussion of related work is given in section 5. The final section of this paper provides a brief summary and outlook.

2 A Taxonomy for Workflow Monitoring and Controlling

2.1 Monitoring and Controlling in the Workflow Life Cycle

Companies face the need to quickly adapt to changing market conditions and customer wishes. Having a transparent overview about ongoing and historical business processes gives companies the flexibility to adjust the treatment of individual cases as well as the opportunity to make structural changes to business processes. Workflow management systems separate application logic from business process logic, enabling users to modify the business processes based on intelligence gathered from the use of a workflow application.

The life cycle of a workflow application has been described by several authors as a closed loop (see e. g. [7]), starting with the definition of the business process to be implemented (1), followed by the transformation of the process model into a workflow model (2). The enactment of the workflow model (run-time) makes up the third phase of the life-cycle (3), whereas the ex-post analysis of executed workflows in the sense of process controlling (4) generates information that is fed back into the process design phase Other workflow cycles, such as the one described by HEILMANN (see. e. g. [9]), contain different phases, but they all depict the development, deployment and analysis of a workflow application as a closed loop.

In practice, the closed loop assumption does not reflect the development and deployment of workflow applications as it actually happens. This is due to the fact that there is no isolated cause-effect relationship between workflow design and enactment. Changes that affect the performance of a workflow can be applied not only to the workflow model itself, but also to the invoked application systems as well as the organizational surroundings of the workflow application. If, for example, the organizational signature power of certain workflow participants is increased, these participants could autonomously approve an increased number of cases, lowering the number of managerial approvals necessary and thus reducing the workload of certain (potentially limited) resources. Even though this measure has no direct effect on the workflow model or the workflow management system, its effects are a noticeable change of the process performance (e. g. lower average throughput time).

In our experience, the introduction of a workflow management system is a sequential process, very similar to the development of a complex application system. Workflow monitoring and controlling are cyclic activities that follow the introduction of the system and run in parallel to system administration and maintenance tasks. Figure 1 shows a procedure model for workflow application development and deployment (cf. [33]). The organizational loop on the right reflects the cyclic auditing of the workflow application. Its feedback is applied in the technical loop through incremental changes to the workflow application. This way, a continuous improvement process can be maintained, without facing the risks associated with a fundamental redesign of existing processes and applications.

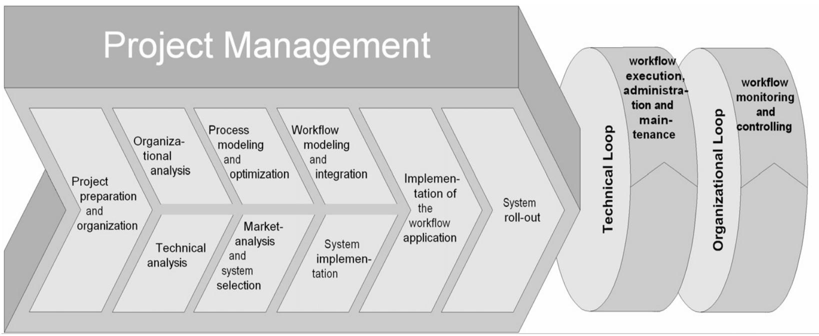


Figure 1: Procedure Model for Workflow Development Projects

2.2 A Taxonomy for Workflow Analysis

The analysis of workflow data can be based on either short-term or long-term observation of the audit trail. This differentiation is typically used to distinguish between workflow *monitoring* (discussed in section 2.3), which describes the analysis of active workflows, and workflow *controlling* (discussed in section 2.4), which deals with the ex-post analysis of (potentially finished) workflow instances. While short- and long-term observation provide a suitable differentiation for the data to be analyzed, the purpose of the analysis can serve as another point of differentiation. With regard to this, technical and business-oriented analysis goals can be distinguished.

From the end user point of view, another separation can be made between the analysis of data for the purposes of an individual user as opposed to the purposes of an organization (for reasons of clarity, this view has been ommitted from the table). These different perspectives determine both the relevant objects for the analysis (activities, resources, data, applications), as well as their cumulation (e. g. the evaluation of single workflow instances as opposed to the evaluation of an aggregate of multiple workflow instances).

Table 1 shows a taxonomy for the analysis of workflow protocol data. For reasons of clarity, the table contains different evaluation opportunities for single workflow instances in contrast to aggregate workflow instances.

The analysis of workflow audit trail data for technical purposes is performed mainly by system administrators and workflow designers, that are debugging workflow models that are not executed correctly (since some workflow management systems use relaxed formalisms for their modeling languages in order to accommodate less formal business process modeling techniques, the resulting processes may contain potential deadlocks or similar problems). Another purpose could be the monitoring of system loads and response times in order to determine the proper scale of the underlying hard- and software-components (mostly in terms of databases). In order to identify workflow instances that are “stuck” on the worklist of an absent user (and the respective activity does not have a “timeout” attribute), the system administrator may query the audit trail for activities that are assigned to absent users. But not only human users take advantage of the logged workflow history. In case of a system failure many workflow systems use the audit trail file to recover the system state to the last committed transaction, much like a database system uses a transaction log file (for a discussion of dabase-related transaction processing and its impact on workflow technology see e. g. [2]).

		Evaluation Focus	
		Technical	Business-oriented
Data Scope	Current Data (live)	Single Instance	Single Instance
		• exception handling (e. g. manual termination of workflow instances, manual reassignment of orphan activities)	• workflow monitoring (e. g. tracking of the individual process state)
		Multiple Instances	Multiple Instances
		• system monitoring (e. g. number of database connections, response time) • license monitoring (e. g. utilized invoked applications)	• workforce management (e. g. scheduling of resources based on predicted workload) • workload management (e. g. assignment of work items based on individual capacity restrictions)
	Historic Data (ex-post)	Single Instance	Single Instance
		• workflow debugging • workflow recovery (in case of a system failure)	• audit purposes (e. g. proof of fulfilment in case of customer complaint)
		Multiple Instances	Multiple Instances
		• license management (e. g. utilized invoked applications)	• workflow controlling (e. g. activity based costing, resource utilization etc.) • workflow planning (e. g. development of new processes and procedures)

Table 1: Taxonomy for workflow audit data evaluation

The business-oriented analysis of workflow audit trail data can be divided into the monitoring of current workflow instances, which is performed by workflow users and process managers (or process administrators), whereas the controlling of past workflow instances is mainly conducted by enterprise controllers. Since these two analysis purposes promise significant business value, we will discuss them in detail in the following sections.

2.3 Process Monitoring

Process monitoring deals with the analysis and overview of process instances at run-time. Using monitoring information, workflow administrators and process managers can adjust the behavior of current workflow instances and react to problems that arise during process enactment. Furthermore, process monitoring is used to improve the responsiveness of an organizations to customer inquiries. When the current state of a process instance can be determined easily, questions such as “Who is handling the customer order 4711?” can be answered in an efficient manner. For the individual workflow participant, monitoring provides the ability to identify those colleagues that worked on a particular case earlier, in case of open issues that need to be resolved. Figure 2 shows the monitoring of an online order while it is being processed by the vendor.

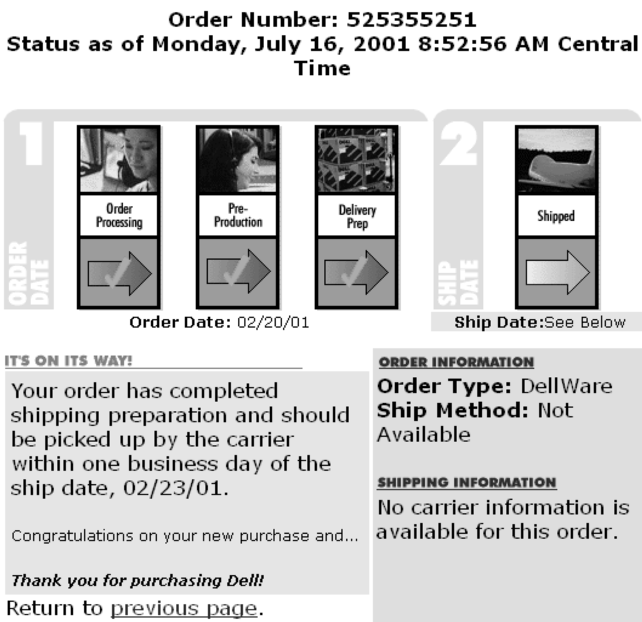


Figure 2: Monitoring of an E-Commerce Order

Process monitoring beyond single process instances can be used to predict staffing requirements. If the average processing times of activities allow for a forecasting of open processes at a specific point in time, the number of active process instances as well as the current activities of these instances allow the short-term prediction of staff requirements. Combined with the ability of workflow management systems to prioritize work items according to the case attributes and the age of the case (i. e. the idle time of a pending case), process monitoring helps companies maintain a consistent level of cycle times even during seasons with high workloads.

The importance of workload transparency is illustrated by an example of a German insurance company. Due to a proposed (and later cancelled) change of the tax legislature, life insurances were subject to additional taxation, if the contract was signed on or after January 1st, 2000. This announcement led to a fourfold increase in life insurance applications during the second half of 1999. The staff at the life insurance department worked overtime to handle the unusual amount of applications, neglecting all other cases that were not new applications. As a result, the structure and age of the remaining cases was unknown and customers complained about the long time it took the insurance to get back to them regarding their inquiries. This situation could have been avoided easily, if a workflow management system had kept track of all the cases and prioritized those, that were older than a certain threshold.

Under certain conditions it is desirable not to expose the detailed process structure to the party monitoring a certain workflow. An example is the presentation of workflow data to workflow participants outside of the company the workflow is executed in, e. g. customers or suppliers. Figure 2 shows the web display of an ordering process at an e-commerce web site. Even though the internal processes are much more complex, only four steps are displayed to the user. This *business state* differs from the actual *process state* in the way that it is a simplified state model of the underlying process state model (NAEF et al call these state models shadow processes [20]). Figure 3 shows an example for the abstraction level between the actual process state and a business state. The four activities in section B and the two activities in section C are combined into the single status B’ and C’, respectively, whereas the activities A, D, and E appear at the same level of granularity in the business state model.

In the (simplified example, the business state model can only contain the same number or fewer states than the process state model, since it is derived from the workflow states exclusively. If context data such as the values of

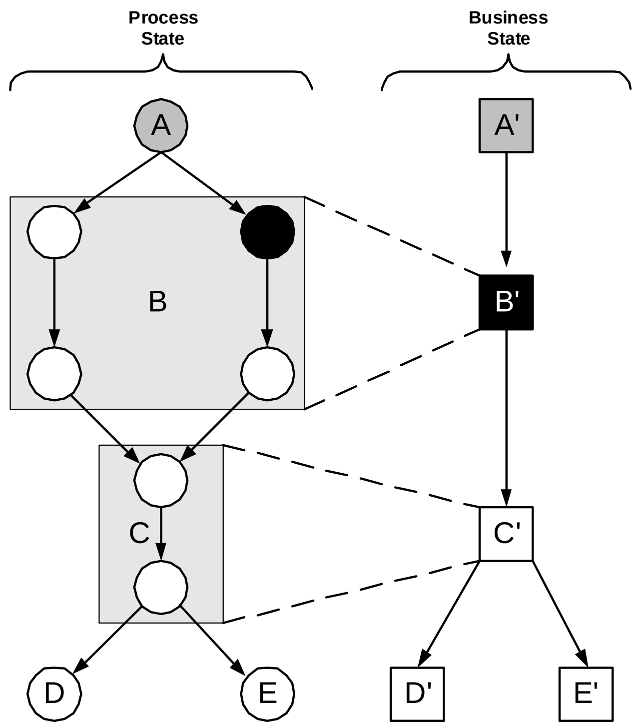


Figure 3: Process State and Business State

certain process relevant variables is taken into account, the coarse states of the business state model may be refined into sub-states.

Besides the organizational process monitoring, workflow management systems typically provide facilities for technical monitoring, which deals with parameters such as response times, system load and the like. With regard to technical monitoring workflow management systems do not differ from complex application systems that are managed through commercial packages such as Tivoli [25] or Candle [4]. Figure 4 shows a screenshot of the technical monitoring facility of a commercial workflow management system. Besides the current numbers of active users, processes and activities the system also displays the number of pending processes and activities (i. e.

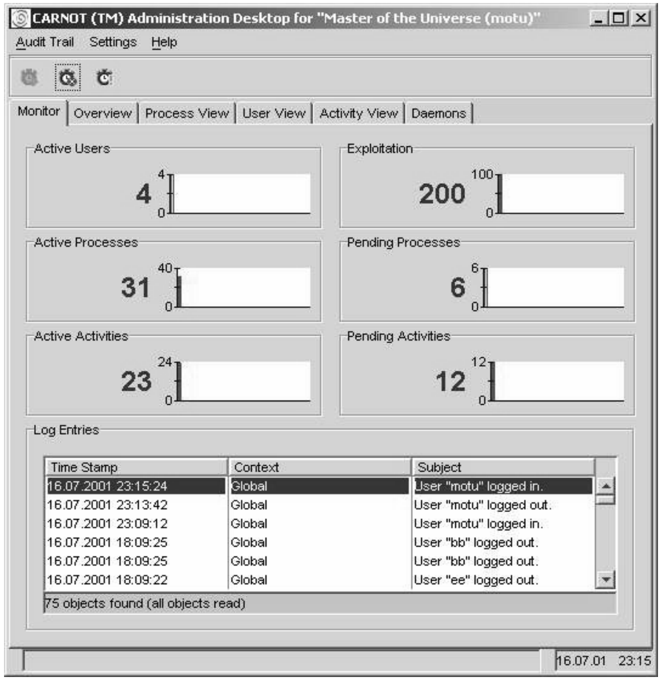


Figure 4: Technical Monitoring Facility

those processes and activities that have been accepted by a user but that have not been completely processed).

2.4 Process Controlling

Process controlling deals with the ex-post analysis of workflow audit trail data. Here the single workflow instances are aggregated according to different evaluation dimensions schemes. Workflow controlling is useful for the detection of long-term developments in workflow enactment and the review of already existing workflow implementations. In order to identify deviations in process execution, the audit trail data is often compared to target data derived from corresponding business process models. The goal of workflow-based controlling is the improvement of future process enactment, thus its effects are more long-lasting than the results of workflow monitoring.

Whereas the target audience for workflow monitoring data is administrative IT personnel on the one hand (for technical information) and workflow participants on the other hand (for organizational information), workflow controlling data is mainly used for enterprise controlling purposes. Analyzed autonomously, audit trail data provides information about the temporal aspects of process execution as well as information about resource utilization on the process and activity level. However, information about the business context of a particular process in most cases cannot be answered by looking at audit trail data alone. This is due to the fact that most workflow management systems do not store data that is processed in the applications invoked during workflow enactment. The WORKFLOW MANAGEMENT COALITION has published a definition of three classes of data associated with workflow systems [30].

- *Application data* is data beyond the control sphere of the workflow management system. It is managed and stored by the applications invoked during the enactment of workflows, e. g. a letter to a customer that is managed by the word processing system.
- *Workflow-relevant data* is managed by applications and has an impact on the control flow of the current process. Typically this type of data is queried at decision nodes during the process, when the workflow management system has to decide which of several alternative paths to follow. Workflow-relevant data can also be used to increase the flexibility of staff assignment rules (e. g. “IF claim.value()<50,000 THEN performer.role() = accountant ELSE performer.role() = manager”). This type of data is read (but not updated) by the workflow management system, but only few systems store this information in their audit trail records.
- *Workflow-internal data* is managed by the workflow management system itself and contains information about the current process instance, e. g. the ID of the process starter or the name of the performer of the last activity. This information is used to realize run-time specific semantics in the process flow, such as the assignment of an activity to the manager of the process starter. This is the kind of information found in most audit trail formats.

Many commercial workflow management systems adhere to the WfMC separation of application data, workflow-relevant data and workflow-internal data. Some systems allow the workflow designer to specify complex data structures and provide facilities for explicit data flow and transformation specifications. Nevertheless, these systems rarely store this kind of data in the audit trail file.

From a controlling perspective, workflow audit trail data is another source of information, just like financial statements from an accounting system or log files from a transaction processing systems. In order to enhance the business value of audit trail data, it needs to be combined with application data. Data warehouses are a common repository for this kind of data, and elaborate on-line analytic processing tools exist that support the controlling recipients during the evaluation of information stored in a data warehouse. Therefore, the integration of workflow audit trail data into the existing data warehouses provides the opportunity of enhancing existing controlling infrastructures with the ability to analyze the business process perspective of an enterprise. This task is non-trivial and we discuss the implications in section 3.

3 Data Warehouse Design for Process Information Systems

3.1 Transforming Audit Trail Data Into A Data Warehouse Schema

During the extraction, transformation and loading phase of data warehouse development, source data is converted into a format that fits the data warehouse schema. Depending on the format of the existing audit trail logs, that could be database records or flat file structures, an appropriate import mechanism needs to be deployed. Using a transformation algorithm, the proprietary schema of the audit trail is converted into the data base schema of the data warehouse. Since the audit trail data is machine-generated, the removal of missing or incorrect values is of little importance during this step.

The audit trail schema used by the workflow management system has to be converted to a schema that is suited for the analytical purposes of the data warehouse. If audit trails from different workflow management systems have to be integrated, a common schema for all workflow audit trail formats involved has to be developed. The WORKFLOW MANAGEMENT COALITION has published a common audit trail format, that could be used as a common denominator by different workflow vendors (WfMC Interface 5 specification [29]). In practice, however, most workflow vendors have defined a proprietary format for their audit data. Out of 20 vendors claiming conformance with WfMC standards, 9 have expressed interest in supporting the Interface 5 standard, but only 2 have actually implemented the standard. It should be noted, that currently no official conformance testing facility for the WfMC standard exists, so even the two implementation have not been evaluated by an independent organization yet.

Figure 5 shows a section of the WfMC Interface 5 data format. Each audit trail entry consists of a CWAD (Common Workflow Audit Data) prefix and suffix. While the prefix is defined by the WfMC, the suffix can be used for vendor-specific extensions. Depending on the type of event that occurred in the system, a different set of attributes is recorded in the body of the log entry. The example shown in the figure depicts an audit trail entry for the creation or start of a process or subprocess. Accordingly, e. g. the termination of an activity would result in a different set of data recorded in the body of the log entry. For the transformation of audit trail records into a common data warehouse schema, this results in the filtering of excess data and the replacement of missing values, that are not recorded for certain event types.

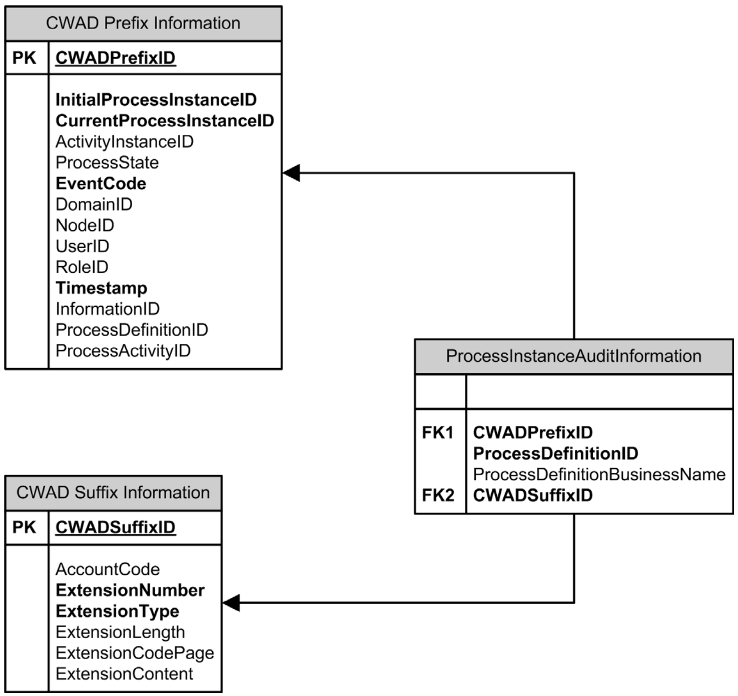


Figure 5: WfMC Audit Trail Format (excerpt)

While the events (such as state changes) during workflow enactment are the point of reference for the recording of the audit trail information, most users of the audit trail will analyze the data at a higher level of integration. Evaluations at the process or activity level require the integration of several audit records. If a workflow system supports multiple state changes between the suspend and resume of an activity or a process instance, the number of audit trail entries for an activity is not predefined. Integrating these entries can resulting in complex join operations on large data volumes. For the design of a process-oriented data warehouse this implies, that during the extraction, transformation and loading phase, summarized datasets have to be generated for activities and processes, to ease subsequent evaluations.

3.2 Integration of Business Object Data

As pointed out in section 2.4, most audit trail formats do not contain a link to the application data that has been processed in the workflow instance. However, meaningful evaluations in a business context almost always require the workflow audit data to be linked to some business object information, such as the customer account involved or the merchandise handled during the process. In our previous research prototype PISA (cf. [32]), we manually added an activity that created a database entry containing the ID of the process instance and the OID of the business object treated in the process. This artificial linking of the audit trail data with (semantic) business data provides the opportunity to navigate from a particular process instance to the business context and allows for the application of numerous evaluation methods.

The linking of audit data with business object information can create a number of problems for the data warehouse designer:

- *Not every workflow context is a business object*. If, during the course of a workflow instance, data in different application systems is used, an artificial wrapper for this particular data set has to be created for every workflow instance in order to make this data set accessible through a unified OID. While object-oriented workflow management systems (e. g. those running inside a java-based application server) provide a native support for this concept, most stand-alone workflow systems are designed to work with different applications without an additional wrapper.

- *Business data is subject to side-effects*. The notion of application data was introduced by the WfMC to make sure, that mission critical data was not locked from use while being used by a (potentially long-running) workflow instance. Reverting this view, application data can be manipulated by applications outside of the workflow context at any time. If audit trail data is transferred through a batch procedure once a day, intermediate changes to the business data are invisible to the data warehouse. To avoid this problem, business data would need to be transferred to the data warehouse as soon as the workflow instance treating this business object is finished. For performance reasons, this will only be feasible in few cases.

- *Synchronicity is not always desired.* From a controlling point of view, it is possible that the most recent version of a business object is of more value than the version that was current at the time of workflow enactment. When, for example, a query is issued "Show all order fulfilment processes from Q1/2000 for customers residing in the U.K." the result set may or may not reflect the customers domicile during process enactment. The correctness of the result set depends very much on the objective of the person analyzing the data. If the purpose of the query was the identification of buying patterns for a direct marketing campaign, subsequent relocations of customers should be included. If, however, the goal of the query is the identification of demographic profiles of specific geographic regions, the customer addresses as of the time when the processes were executed is relevant. This example shows, that the purpose of the data recipient determines the synchronicity requirements for audit trail and business data during the import phase.

- *Recording audit trails creates potentially large amounts of data.* LEYMANN and ROLLER have pointed out the risk of growing audit trails [14]. They estimate the typical number of log entries per activity to be 5 (or 1 KB for a fully featured audit trail entry). For 10,000 processes with 10 activities per day this creates an audit trail of 500 MB every day, not including the associated business object information. Numbers like these are not unusual. The average number of pages received in the mail room of the insurance company described in section 4 is on average 29,000 per day, resulting in 8,900 different cases that are instantiated every day.

- *Processes evolve*. The opportunity to model and change processes with a workflow management system is significantly easier than the adjustment of hard-coded process logic in application systems. Therefore, over time, enterprises using workflow applications will adjust these to match their changing environment. The change of the process structure, e. g. the replacement of an activity with another results in a change of the audit trail data during process enactment. If an analysis is to be performed across a collection of process instances, in many cases only those instances will be relevant, whose execution path is identical, i. e. whose set of activity instances is identical at the type level. Therefore, it is vital for the validity of the data warehouse information, that for every workflow instance the underlying workflow model variant is recorded.

3.3 Evaluation Dimensions

During the evaluation of workflow audit trail data, different evaluation dimensions are used and various evaluation methods can be executed on the data repository. While not all evaluation requirements of the data warehouse users may be known in advance [8], some basic evaluation perspectives can be predefined. In its most basic format, workflow audit trail data supports analyses from a process perspective as well as from a resource perspective.

3.3.1 Process perspective

The process evaluation perspective focuses on information that can be derived from an analysis of the temporal and logical information associated with the activities and transitions recorded in the audit trail. Without any modification, this information is derived from the activity identifiers as well as the timestamps and status recorded in the audit trail.

Typically, a workflow audit trail contains the following timestamps:

- Instantiation time of the process
- Starting time of the process
- For each activity
    - creation time (the time when the activity becomes ready for execution)
    - activation time (the time when the activity was activated by a workflow participant, e. g. by selection from the participant's work list)

- completion time (the time when the activity was marked as completed either by the system or the participant in case of a manual activity)

- Completion time of the process

Using these timestamps and the differences between them the following information can be derived easily:

- Cycle time of the process (the time elapsed between the instantiation or starting time and the completion of the process)
- Net processing time (the sum of the time elapsed between the activation and completion of each activity of the process)
- Idle time (the time elapsed between the creation time and activation time of each activity).
- Frequency of process occurrences (which processes were executed in which frequency at a certain hour/day/week/month etc.)

Figure 6 shows different analysis scopes, namely at the activity, segment and process level. The resulting analyses can be performed at the instance level (for the analysis of the behavior of single activities and processes) as well as at an aggregated level (for the analysis of the behavior of activities and processes over time or the computation of average values).

A frequency analysis can provide information about the likelihood of the execution of certain activities and/or process paths. This information is of purely statistical value, if no connection can be made to the business objects that were treated in the different process instances (i. e. the reason, *why* a process instance was executed in a certain fashion). This does not mean, however, that these results are not meaningful at all. For example, probabilities about the execution of certain process paths could be inserted into the workflow model for simulation purposes, enabling the process designer to experiment with different workloads or staff assignments before actual changes are made to the organization.

In combination with the semantics of the process and the business objects involved, the results of a frequency analysis can be used to identify potential weaknesses or bottlenecks along the process. For example, if the variance of the cycle time for an order fulfilment process is particularly large, the analysis of the business objects associated with extremely fast (or slow) processes may yield some information about the reason for this non-homogeneous behavior.

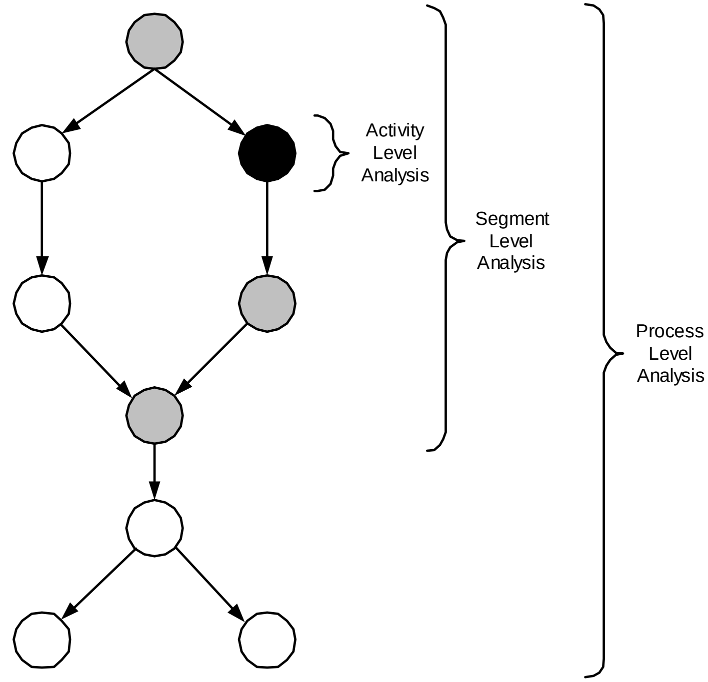


Figure 6: Analysis at the activity, segment and process level

3.3.2 Resource perspective

Most workflow management systems allow for the recording of the performer ID in the audit trail. In our experience, during the introduction of a workflow management system, this technical option often creates mistrust and resistance among the future users of the system. The measurement of individual performance through an information system is subject to privacy laws in some countries. In Germany, for example, the measurement of employee performance through information technology requires the agreement of the employee’s union. If no employee’s representatives exist at the company in question, a visible notice has to be posted at the workplace under surveillance. Therefore, a sensible measure during the initial setup of a workflow management system is the creation of a formal charter that contains the level of detail for the audit trail as well as the evaluation purposes, this data is collected for.

Combining temporal information about process performance with information about the resources involved provides opportunities to create meaningful financial indicators. If the cost factor for the utilization of a resource is known, the (personnel) cost for the execution of a process instance an be computed by multiplying the processing times of the activity instances with the cost factor of the workflow participants. This simplified approach does not take into account work that is performed outside the scope of the workflow management system, e. g. answering to a customer call while another customer order is being processed inside the workflow system. However, for organizations that use workflow systems for administrative processes (where personnel cost account for the majority of process cost), the use of workflow audit trail data can lead to a better accountability of overhead cost.

Another evaluation using the resource perspective is the analysis of resource utilization at the activity or process level. This analysis answers questions such as "How many different participants were involved in an ordering process?" or "How many work items typically reside in the work list of an accounting clerk on average?"

In [32] we have discussed the application of SASSONE'S hedonic wage model to workflow audit trail data. By classifying workflow participants into groups such as executives, experts and administrative staff, and by assigning categories to workflow activities that reflect these groups (for example executive tasks, administrative tasks, expert work), the analysis of audit trail data provides information about the accuracy of work assignments. Using the hedonic wage model, the financial effects of organizational restructuring on process performance can be visualized.

3.3.3 Adding Perspectives through Business Objects

The combination of audit trail data with semantic information about the business objects allows for a multitude of evaluations. Depending on the process semantics as well as the information available about the business objects, numerous evaluation dimensions can be derived.

An example for an analysis originating at the business object perspective is the analysis of processing times grouped by the vendor of the process object. For a supply chain process, this information could be used to classify suppliers into A, B and C suppliers, according to the frequency of exceptions generated within the processes triggered by the different companies. Using a combination of business object and audit trail data, an analyst can navigate along the process structure as well as the data structure of the associated business object. Starting from the process perspective ("what is the average cycle time of an invoice auditing process?") the analyst could single out processes according to their temporal behavior ("display those process instances, whose cycle time exceeded the average cycle time by more than a week"). Using this information, the analyst could traverse to either the resource or the business object perspective ("out of this set, display the name of the supplier, whose invoices exceeded the cycle time the most"). From an enterprise controlling perspective, these are the most valuable insights, the use of workflow audit trail data can generate.

The number and kind of evaluations available at the business object perspective is determined mainly by the semantics of the business objects. Ideally, the evaluations are defined at the time, the workflow application is being implemented, to ensure that the data required for the analyses is recorded (either through the workflow management system directly, or through external programs).

3.4 Designing a Process-Oriented Data Warehouse

Figure 7 shows the schematic design of a data warehouse that integrates workflow audit trail data with business object information. The extraction, transformation and loading layer imports workflow audit trail data as well as business object data into a common repository. The data is formatted according to the respective metadata sets that describe the format and semantics of the underlying raw data structures.

The evaluation engine performs various evaluation methods on the process-oriented data warehouse and delivers the results as an evaluation result set. For performance and warehouse size reasons it may be feasible to create data marts for specific processes or analysis purposes (e. g. a specific data mart for activity-based costing analyses). Finally, the visualization layer reads the evaluation results and presents them to the user in a usable format (e. g. as a table or a graphical chart, depending on the type of result data).

The arrows on the left and right of the data warehouse diagram indicate the steps taken during design and use of the process-oriented data ware-

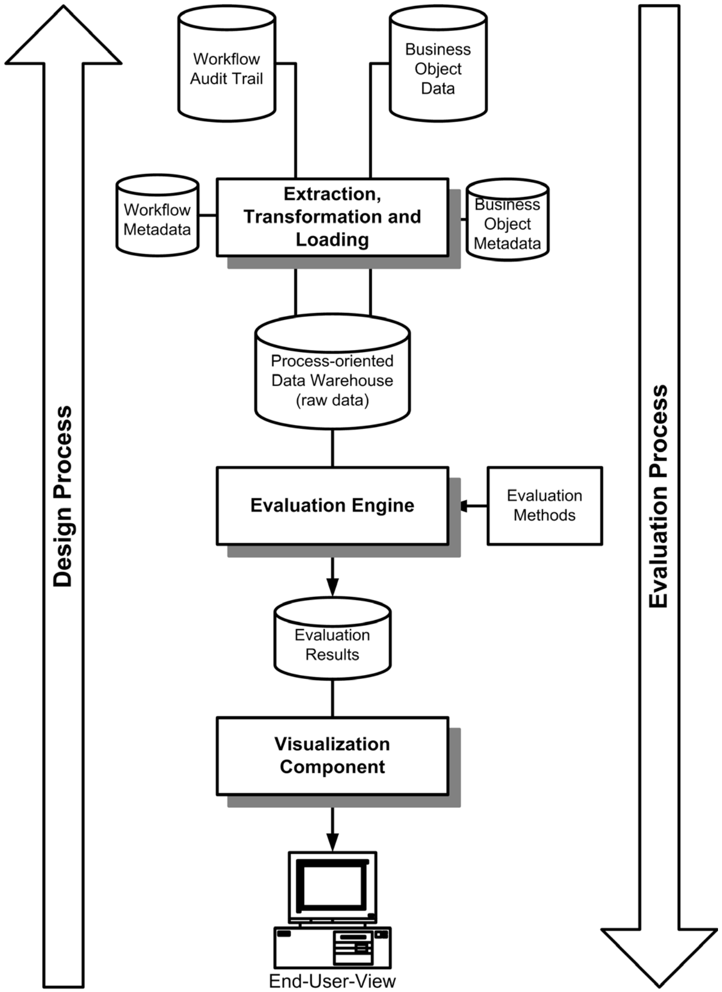


Figure 7: Design of a Process-oriented Data Warehouse

house. During the operation of the process-oriented data warehouse, a continuous feed of audit trail data will be integrated (either continuously or via batch procedures), evaluated and presented to the user. The design process reverses this flow. Starting with the business requirements of the user (what kind of business problem or decision needs to be addressed by the user?) the information necessary for the fulfilment of these requirements can be identified (what does the user need to know in order to solve the problem at hand?). The information requirements are the major guideline for the selection of an appropriate evaluation method (which evaluation delivers the results needed?). Finally, the choice of an evaluation method determines the required entries in the process-oriented data warehouse (what source data is required for the execution of this evaluation method?).

This top-down approach at the design of a process-oriented data warehouse illustrates, that the managerial controlling requirements need to be stated already at the time the workflow is designed. Subsequent changes to the evaluations required may result in necessary adjustments of the underlying workflow application, which can be both time-consuming and expensive.

4 Case Study: Using Workflow Audit Trail Data in an Insurance Scenario

4.1 Preconditions

In the following section we outline the practical relevance of our approach through a case study. The study was conducted at a medium-sized German insurance company with 1,200 employees. The company had conducted a large business process re-engineering project two years before the case study. During the course of this project, the major processes affecting the insurance-specific departments had been documented and grouped into clusters. Based on statements from employees, the average processing time for activities was documented (with a resolution of minutes). Based on this information, an activity-based costing system (ABC) had been implemented.

After the new system had been in use for two years, it became apparent that the underlying business processes had changed. Also, certain information that was required for purposes such as capacity planning was not available from the existing system. A project was set up to investigate the potential benefits of a workflow-driven process information system for the company.

4.2 Information Availability

Figure 8 illustrates the existing activity-based costing system. It is characterized by a variety of data feeds, most of which are transferred manually into the activity-based costing system. The activity structure of the business processes was transferred manually from Excel sheets into the proprietary ABC-system. Transactions from the operational insurance application system were assigned to activities according to a transformation schema developed during the BPR project. Since the number of transactions were recorded in the legacy system, the log files were used to compute the actual number of activities performed in a certain period. Using the estimated processing times per activity, the required personnel capacity per period was computed using a custom Excel application. This data was combined with the activity structure and the internal accounting information, which was collected in a data warehouse application and manually transferred into the ABC-system.

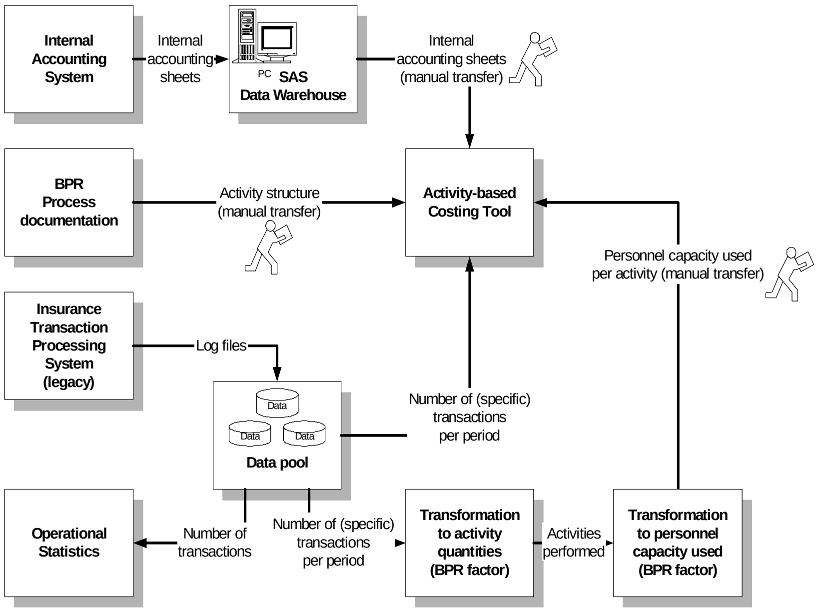


Figure 8: Controlling Data Flow (As-Is)

Besides the fact that the whole evaluation process was time-consuming and error-prone, due to many media-breaks and manual transfers, the results of the ABC-system did not reflect the actual operations at the insurance company. For example, only those processes appeared in the ABC-system, that resulted in the execution of a (traceable) transaction in the transaction processing system. This way, the ABC-system documented 20,000 cases of customers cancelling their auto insurance per year. However, the actual number of cancellations was much higher, since many customers could be convinced to revoke their cancellation request, before it was put into effect using the transaction processing system. These cases, even though their processing consumed time and kept resources, did not appear in the ABC-system. Furthermore, changes to the transaction processing system jeopardized the link between the BPR activity structure and the transactions, since the measured amount of time necessary to perform a certain activity was tied to the transaction used within the activity. When a transaction was changed the processing time for the associated activity changed, but this change was not documented anywhere.

4.3 Improvement through Workflow Data Integration

After the identification of these shortcomings, the concept for a workflow-based data warehouse was designed, that should be used as a data feed for the existing ABC-tool. Figure 9 shows the structure of the target scenario. The central component of the future controlling infrastructure is a data warehouse, that serves as the central repository for all operational data the needs to be evaluated. The workflow audit trail is transferred into the data warehouse using two separate import filters. On the one hand, the process and activity information is extracted for the process perspective evaluations. On the other hand, resource information (which resource performed which activity) is extracted for resource perspective evaluations. In order to provide semantic information about the process structure, to-be information about the processes and activities is transferred into the data warehouse at build-time. Manual activities (for example phone calls), however, will still not be recorded by the new system.

In order to link the audit trail information to operational business data, two tables need to be maintained separately. One table contains the relationship between the process instances and the insurance contracts or the customers involved. Using this table as a bridge, a drill-down from a single process to the customer who triggered the process can be performed. The other custom table contains the relationship between the workflow participants and the internal cost accounting structure of the enterprise. Since workflow management systems only record a technical ID of the performers

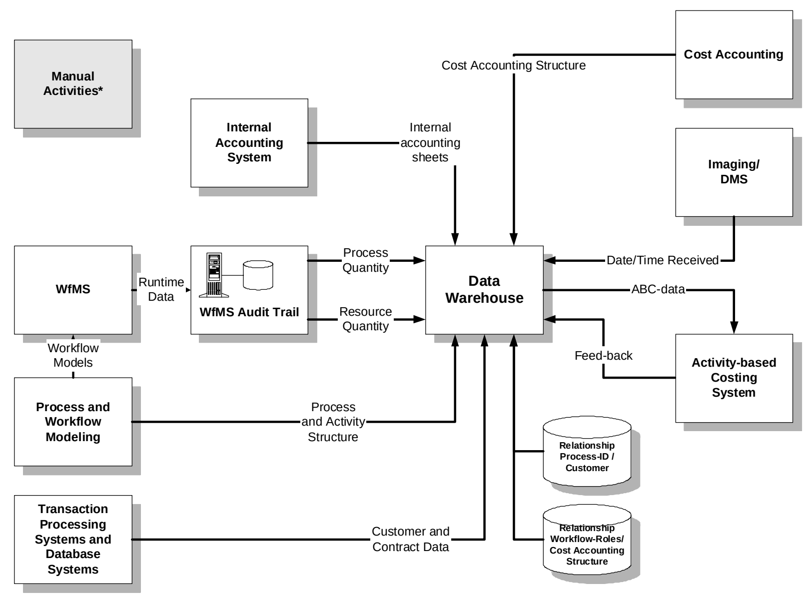


Figure 9: Controlling Data Flow (To-Be)

who executed activities, this table is required to allocate cost factors to the workflow participants.

The realization of the process-oriented data warehouse will enable the insurance company to evaluate its operational business processes in a timely fashion and with a much higher level of accuracy than currently possible. It should be noted, that the existence of a workflow infrastructure is the major precondition for the realization of such an information system. At the company in question, a project was initiated to lay the foundation for a future data warehouse by selecting and implementing a combined document management and workflow infrastructure. Even though the financial benefits of an improved process controlling infrastructure are hard to determine, the financial savings from the introduction of the document management and workflow infrastructure were sufficient to provide a return on investment just three years after the deployment of the new system. The ability to perform detailed process analyses is perceived as an added benefit of the new infrastructure and work packages for the realization of the proposed infrastructure have been added to the project plan.

5 Discussion

There are relatively few research projects that aim at combining data warehousing concepts with workflow management systems. The existing sources can be grouped into two categories: Those who apply workflow concepts to the data warehouse domain and those who describe the analysis of workflow audit trail data.

5.1 Applying Workflow Concepts to Data Warehouse Design

Sources that use workflow technology for the design of data warehouses aim mainly at the design phase of a data warehouse. A notable exception is the work by LIST ET AL. who aim at designing a generic data warehouse for business process evaluation [15]. Since the retrieval and transformation of operational data into the data warehouse is a frequent process, whose steps are structured, it is a potential candidate for workflow automation.

BOUZEGHOUB et al. describe the modeling of the data warehouse refreshment process as a workflow [3]. They use an event-driven approach to workflow modeling to trigger various parts of the refreshment process either after a certain timer has expired or after a certain condition in the data warehouse has occurred. In particular, they distinguish between *Client-driven refreshment*, when a user activity causes an update to the data warehouse structure, *Source-driven refreshment*, when the changes to the source data of the data warehouses trigger a refreshment process, and *ODS-driven refreshment*,

PATTERSON discusses workflow and data warehouse concepts and identifies several areas in the data warehouse design process that could benefit from the use of workflow technology, such as handling update anomaly problems [21].

5.2 Analyzing Workflow Audit Trail Data

The application of data warehouse concepts to workflow technology can be divided into three groups. The first group of publications in this area deals with the technical facilities necessary for the logging of the audit trail data (e. g. [27]). A second group of papers describes the reverse engineering of workflow processes from audit trail data [1]. Finally, a number papers can be identified, that focus on the ex-post analysis of workflow audit trail data with methods known from enterprise controlling ([5],[22],[19],[28],[32],[33]).

5.2.1 Technical Facilities for Audit Trail Generation

KOKSAL et al.discuss the management of audit trail data in the distributed workflow management system Mariflow [11]. The paper focuses on technical issues regarding the storage of audit data and economic queries on distributed data sources, but the authors do not address the business value of the history data.

An approach for the tracking of history information in a distributed workflow management system is presented by MUTH et al. [18]. Within the prototype Mentor-lite, data about current and past workflow instances are kept in a temporal database that can be queried either at runtime or for ex-post analyses.

5.2.2 Reverse Engineering Using Audit Trail Data

AGRAWAL et al. use data mining techniques to create workflow models from audit trail data [1]. The purpose of their project is to use data from the ad-hoc execution of processes, to subsequently identify common rules and procedures and to create workflow models using a bottom-up approach. The authors have tested their approach against artificial data sets and audit trail data from a live workflow installation. However, the practical use of the methodology presented requires the existence of a flexible workflow tool, that records processes while they are being created on the fly. Despite the obvious demand for such a tool, the current workflow market is lacking products with this kind of execution flexibility.

5.2.3 Workflow-based Controlling in the Literature

Whereas the controlling of processes using workflow audit trail data has been analyzed in the German literature to some extent (for example [5],[22],[28]), there are relatively few English sources dealing with this topic (for example [19],[32],[33]).

The WORKFLOW MANAGEMENT COALITION Interface 5 specifies the elementary information a workflow management system should record the execution of workflow instances [29]. The existing standard provides a data format for the audit trail data as well as guidelines, which events should be recorded. However, the evaluation of this information is not addressed in the WfMC standard. After the publication of the current standard document, a proposal for an API has been submitted to the Coalition, but so far no progress has been made to change the interface from a data format specification to a functional specification.

MCGREGOR discusses the existing WfMC standard with regard to a prototype, that uses this information to generate the process perspective of a balanced scorecard [16]. Her work aims at the development of a closed-loop system that takes audit trail information as an input and delivers advice to process designers, which aspects of the workflows analyzed could be optimized.

One of the first sources on workflow based controlling can be found in [19]. MCLELLAN provides an overview of the analysis of historical process data. He discusses the evaluation of workflow history data as workflow metrics. The controlling applications described are statistical evaluations as well as the run-time detection of late cases and overdue tasks.

SHOHAIEP ET AL. present a methodology for the identification of process knowledge through the analysis of work practices [24]. Even though their method is called “Workflow*BPR”, it does not require the use of a workflow management system. Taking the basic idea of determining the knowledge incorporated in a process, workflow audit trail data can be perceived as a “knowledge store” in its own right. For example, the experience of a workflow participant could be computed taking the number of times into account that this participant has performed a certain activity type. When this information is fed back into the workflow management system, new staff allocation methods could be implemented, such as “assign this activity to the most experienced person available”.

DERSZTELER discusses the design of a controlling tool called WorkFlow Analyzer in [5]. A partially functional prototype was implemented on the basis of Forest&Trees. Based on audit trail data from the workflow management system WorkParty by Siemens Nixdorf and to-be data from the business process modeling tools ARIS and Bonapart, the prototype provides several quantitative evaluation methods. However, the combination of the audit trail information with business data was not realized using this approach.

In [22], RAUFER discusses the controlling of workflow-based processes using a case study from the manufacturing domain. In his example, the author focuses on a specific process, which he enhances with cost information as well as target work- and cycle-times. The presented prototype, which is based on the workflow management system COI, is targeted specifically at the process analyzed by the author. As a result, the system architecture is not easily generalizeable.

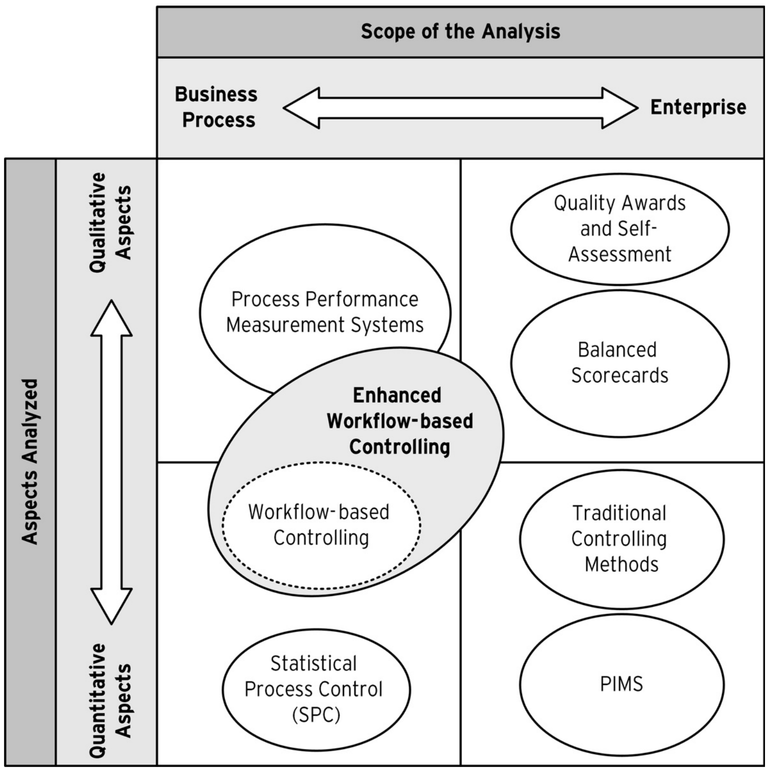
A similar prototype is presented by WEISS in [28]. He developed a workflow-driven activity-based costing system for the commercial workflow management system Staffware. The focus of this approach is the realization of a single evaluation method, therefore, the resulting system is not designed to be extended by additional evaluation methods.

The COPPA project (Computer-based Process Performance Measurement) conducted at the University of Fribourg, Switzerland, deals with the design of a performance measurement system ([12], [13]). Using a three stage approach, the authors first surveyed the market and corporate practice of performance measurement. During a second stage the architectural and functional requirements of a performance measurement system were outlined, and during a third phase, a prototype of the performance measurement system was implemented. In relation to workflow-based controlling, the authors position process performance measurement at a higher level of abstraction, that includes information about the strategic positioning of an enterprise, whereas a workflow-based controlling system is mainly focused on the analysis of operational data.

6 Summary and Outlook

Workflow-based controlling provides companies with the ability to accurately measure the operational performance of business processes. Combined with data warehouse technology and operational business data, complex evaluations can be performed, that help enterprises to assess their current situation more precisely than the sole use of traditional key performance indicators. The increasing maturity of workflow and data warehouse products provide an important precondition for the implementation of workflow-driven controlling systems.

It should be noted that workflow-based controlling does not replace other controlling mechanisms of the enterprise, but enhances them significantly. Figure 10 shows the positioning of workflow-based controlling in relation to other controlling techniques. While strategic controlling instruments like the Balanced Scorecard focus on qualitative aspects of the enterprise, and traditional controlling methods rely on financial information, few techniques exist, that put the business process into the center of attention. Workflow-based controlling tools that focus entirely on information contained in the audit trail are limited to the quantitative analysis of the event-driven history of process execution.



Adapted from [12], pp. 425 ff.

Figure 10: Positioning of Workflow-based Controlling

Our research enhances workflow-based controlling approaches to include business information that is not stored in the traditional audit trail. This way, the scope of potential evaluations is widened to include qualitative information, and the differentiation of business processes through their associated business objects allows for a more detailed segmentation of the analysis domain. Since this approach relies on the workflow-enabled business processes of an enterprise, there will be a number of blind spots at the enterprise level (i. e. manual processes, managerial processes without workflow support etc.). However, the development of enhanced workflow-based controlling systems can serve as a data-feed to other controlling approaches, like the Process Performance Measurement Systems discussed by KUENG in [12] or the process perspective of the Balanced Scorecard.

We have discussed the technical and organizational aspects of building a process-oriented data warehouse and used a case study to illustrate the practical relevance of this type of application. Nevertheless, many aspects of workflow-based controlling systems require further research.

So far, existing papers focus on the realization of single evaluation methods, like activity-based costing [28]. An analysis, which existing evaluation methods can be applied in the context of a workflow-based controlling system based on the attributes recorded in workflow audit trail data would provide further insight about the coupling of audit trail data and business object information.

Also, the handling of changing process models variants over time resembles the problem of schema evolution in databases. Further research in this direction would help to open workflow-based controlling systems to incorporate audit trails from workflow management systems that support ad-hoc processes.

In the context of the insurance project described in section 4, we are currently working on a prototypical implementation of the concepts presented in this paper using a java-based workflow management system which will be evaluated using operational data from the business processes of this company.

7 References

[1] Agrawal, R., Gunopulos, D., Leymann, F., “Mining Process Models from Workflow Logs”, in: Advances in Database Technology - Proceedings of the 6th International Conference on Extending Database Technology, ed. Fèlix Saltor, Isidro Ramos and Gustovo Alonso, Valencia, 23-27 (1998).

[2] Alonso, G., Agarwal, D., Abbadi, A. E., Kamath, M. and Gunthor, R., “Advanced transaction models in workflow contexts.” in: Proceedings of the 12th IEEE Intl. Conference on Data Engineering, New Orleans, LA, February (1996).

[3] Bouzeghoub, M, Fabret, F. and Matulovic-Broqué, M., “Modeling Data Warehouse Refreshment Process as a Workflow Application”, in: Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW ’99), eds. Stella Gatziu, Manfred A. Jeusfeld, Martin Staudt, Yannis Vassiliou, Heidelberg, 6.1-6.12 (1999).

[4] Candle Corp.: www.candle.com

[5] Derszteler, G., “Prozessmanagement auf Basis von Workflow-Systemen”, Josef Eul Verlag, Lohmar et al. (2000).

[6] Emery, P., “The Workflow Market: A Global Perspective for 2000”, in: Excellence in Practice, Volume IV, Innovation and Excellence in Workflow and Knowledge Management, ed. Layna Fischer, Future Strategies, Lighthouse Point, 41-47 (2000).

[7] Galler, J. and Scheer, A.-W., “Workflow-Projekte: Vom Geschaeftsprozessmodell zur unternehmensspezifischen Workflow-Anwendung”, in: Information Management, Vol. 10 No. 1 (1995), 20-28.

[8] Gardner, S. R., “Building the Data Warehouse”, in: Communications of the ACM, Vol. 41, No. 9, (1998), 52-60.

[9] Heilmann, H.: “Die Integration der Aufbauorganisation in Workflow-Management-Systeme” in: Information Engineering, ed.: Heidi Heilmann et al.. Oldenbourg, Munich, Vienna 147-165 (1996).

[10] Inmon, W. H., Imhoff, C. and Sousa, R., “Corporate Information Factory”, 2nd Edition, Wiley, New York et. al., (2001).

[11] Koksal, P., Arpinar, S. N. and Dogac, A., “Workflow History Management”, in: Sigmod Record 27 (1998) 1, 326-348 (1998).

[12] Kueng, P., “Supporting BPR though a Process Performance Measurement System”, in: Business Information Technology Management, ed. P. Banerjee et al., Har-Anand Publications, New Delhi, 422-434 (1998).

[13] Kueng, P., “Process Performance Measurement System: a tool to support process-based organizations”, in: Total Quality Management, Vol. 11, No. 1 (January 2000), 67-86 (2000).

[14] Leymann, F. and Roller, D., “Production Workflow - Concepts and Techniques”, Prentice Hall, Upper Saddle River (2000).

[15] List, B., Schiefer, J., Tjoa A. M. and Quirchmayr, G., “The Process Warehouse - A Data Warehouse Approach for Business Process Man-

agement", in: e-Business and Intelligent Management - Proceedings of the International Conference on Management of Information and Communication Technology (MiCT 1999), Copenhagen, Denmark (1999).
[16] McGregor, C., "The Impact of Business Performance Monitoring on WfMC Standards", presentation at the Workflow Management Coalition meeting, New York, NY, May 2001.
[17] Moore, C., "Workflow Goes Mainstream", Giga Information Group Planning Assumption, Cambridge, MA, April 21, 2000.
[18] Muth, P., Weissenfels, J., Gillmann, M. and Weikum, G., "Workflow History Management in Virtual Enterprises Using a Lightweight Workflow Management System", in: Proceedings of the Ninth International Workshop on Research Issues on Data Engineering, Sydney, 148-155 (1999).
[19] McLellan, M.: "Workflow Metrics - One of the great benefits of workflow management" in: Praxis des Workflow-Management, eds. Hubert Oesterle and Petra Vogler, Braunschweig, 301-318 (1996).
[20] Naef, A., Schuler, Ch. and Schuldt. H., "Monitoring komplexer Dienste in unternehmensübergreifenden Prozessen am Beispiel von SAP R/3 Business Workflows", in: Proceedings of the 9th GI-Conference Datenbanksysteme in Buero, Technik und Wissenschaft (BTW 2001), 85-94 (2001)
[21] Patterson, J. E., "Workflow and Data Warehousing Taxonomy", Masters Thesis, University of Texas, Arlington (1996).
[22] Raufer, H., "Dokumentenorientierte Modellierung und Controlling von Geschäftsprozessen", Deutscher Universitaets-Verlag, Wiesbaden (1997).
[23] Sassone, P. G., "Cost-benefit methodology for office systems", in: ACM Transactions on Information Systems, Vol. 5, No. 3, (1987), 273-289.
[24] Shohaiep, G., Housel, T. J. and Kanevsky, V.: "Workflow-BPR White Paper: Business Process Auditing with Workflow-BPR", White Paper, University of Southern California, (1997). URL: http://www-rcf.usc.edu/~housel/IOM499/wp.html [2001-07-15].
[25] Tivoli Corp.: www.tivoli.com
[26] Vassiliadis, P., Quix, C., Vassiliou, Y. and Jarke, M., "A Model for Data Warehouse Operational Processes", in: Proceedings of the 12th International Conference on Advanced Information Systems Engineering (CAiSE 2000), eds. Bengt Wangler and Lars Bergman, Stockholm, 446-461 (2000).
[27] Weikum, G.: "Workflow Monitoring: Queries on Logs or Temporal Databases?" Position Paper at HPTS'95. Pacific Grove, (1995).
[28] Weiss, D., "Prozesskostenrechnung und Workflow Management", Deutscher Universitaets-Verlag, Wiesbaden (1998).
[29] Workflow Management Coalition, "Audit Data Specification. Version 2", Document Number WFMC-TC-1015, Winchester (1999).
[30] Workflow Management Coalition, "Terminology and Glossary" 3rd Edition, Document Number WFMC-TC-1011, Winchester (1999).
[31] zur Muehlen, M. and Allen, R., "Embedded vs. Autonomous Workflow - Putting Paradigms into Perspective", in: Excellence in Practice, Volume IV, Innovation and Excellence in Workflow and Knowledge Management, ed. Layna Fischer, Future Strategies, Lighthouse Point, 49-58 (2000).
[32] zur Muehlen, M. and Rosemann, M., "Workflow-based Controlling", in: Proceedings of the 34th Hawai'i International Conference on Social Sciences, ed. Raplph Sprague Jr., Wailea (2000).
[33] zur Muehlen, M., "Workflow-based Process Controlling - Or: What You Can Measure You Can Control", in: Workflow Handbook 2001, ed. Layna Fischer, Future Strategies, Lighthouse Point, 61-78 (2000).