Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
International Journal of Information Management 25 (2005) 458–472
International Journal of Information Management
www.elsevier.com/locate/ijinfomgt

The effectiveness of workflow management systems: Predictions and lessons learned

Hajo A. Reijers[a,*], Wil M.P. van der Aalst[b]

[a]Department of Technology Management, Eindhoven University of Technology, PO Box 513, Pav. D14, NL-5600 MB, Eindhoven, The Netherlands
[b]Department of Technology Management, Eindhoven University of Technology, PO Box 513, Pav. D02, NL-5600 MB, Eindhoven, The Netherlands

Abstract

Workflow management systems are widely used and reputable to improve organizational performance. The extent of this effect in practice, however, is not investigated in a quantitative, systematic manner. In this paper, the preliminary results are reported from a longitudinal, multi-case study into the effectiveness of workflow management technology. Business process improvement is measured in terms of lead time, service time, wait time, and resource utilization. Significant improvement of these parameters is predicted for almost all of the 16 investigated business processes from the six Dutch organizations participating in this study. In addition, this paper includes lessons learned with respect to the simulation of administrative business processes, data gathering for performance measurement, the nature of administrative business processes, and workflow management implementation projects.

Keywords: Workflow management; Business processes; IT effectiveness; Simulation

1. Introduction

Commercial workflow management (WfM) systems have been around since the early nineties, while their conceptual predecessors range back even further, see e.g. Ellis (1979). They have

*Corresponding author. Tel.: +31 0 40 2473629; fax +31 0 40 2432612.
E-mail addresses: h.a.reijers@tm.tue.nl (H.A. Reijers), w.m.p.v.d.aalst@tm.tue.nl (W.M.P. van der Aalst).

0268-4012/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijinfomgt.2005.06.008

become "one of the most successful genres of systems supporting cooperative working" (Dourish, 2001). The worldwide WfM market, estimated at $213.6 million in 2002, is expected to redouble by 2008 (Wintergreen, 2003). Furthermore, WfM functionality has been embedded by many other contemporary systems, such as ERP, CRM, and call-center software. WfM technology, in other words, has become quite successful and widespread.

The alleged advantages of WfM systems are clear. By having a dedicated automated system in place for the logistic management of a business process, such processes could theoretically be executed faster and more efficiently (Lawrence, 1997). Yet, very little is known about the *extent* of performance improvement an organization may experience in practice. Single case studies are available, e.g. Goebl, Messner, & Swarzer (2001); and Prinz & Kolvenbach, (1996), but do not easily lend themselves for generalization. Few empirical studies that include multiple implementations are known to us. What is more, their focus is not on performance issues, but on aspects such as the appreciation of the technology by end-users (Kueng, 2000), implementation (Parkes, 2002), or the metamorphosis of the project objectives (Herrmann & Hoffmann, 2003). The study most related to our research is that of Oba, Onada, & Komoda (2000), who developed a regression model on the basis of 20 cases to predict the reduction of lead time as a result of WfM implementation. Other available data on performance improvement comes from WfM vendors, who are perhaps not completely unbiased. A study among 100 clients of Staffware, one of the world's largest WfM vendors, indicates for instance that 62.5% of their clients sees increased efficiency as a result of WfM implementation (Staffware, 2000). Unfortunately, this outcome is not accompanied by indications how this figure is established, how the notion of efficiency is made operational or how much efficiency gains are achieved.

The lack of information on performance improvement through WfM system is awkward. Despite the large number of research papers on the subject of WfM, the research community has not been able to express general statements on this subject. This paper is an interim report on a longitudinal, multi-case study (Yin, 1994) into the effectiveness of WfM technology. Its aim is to quantify the contribution of WfM technology to improved business process performance with respect to lead time, wait time, service time, and utilization of resources. In this way, it is an extension of the scope of Oba et al. (2000). Its findings may be of relevance for both workflow researchers and practitioners.

Our study, which is conducted in the Netherlands, is a joint effort by Eindhoven University and Deloitte Management and ICT Consultants. It started in 2001 and is planned to continue until 2007. So far, six organizations are involved who are in the process of implementing WfM technology to support 16 different business processes. All organizations are administrative organizations, both commercial and not-for-profit, ranging from medium-sized to large.

This paper presents the first half of our longitudinal study. Based on actual measurements of the process before the introduction of the WfM system and detailed simulations, we present our expectations on performance improvement for each involved business process for each of the previously mentioned performance indicators. These expectations serve as a prediction for the effectiveness of WfM technology, which can be validated when the implementation has been completed and the WfM-enabled processes are taken into operation. Also, the execution of this study has confronted us with a number of issues that seem worthwhile to communicate. First of all, the use of discrete event simulation to realistically capture the dynamics of administrative business process turned out to be far more difficult than expected. Secondly, we observed characteristics of the administrative business processes themselves which conflicted with both our intuition and findings in literature. The same holds for the course that the projects of WfM implementation took with respect to Business Process Redesign (BPR), which—surprisingly—was hardly applied. For all these issues, we present the lessons we learned. A final objective of this paper is to present our research methodology and generate feedback from the research community.

The structure of this paper is as follows. In Section 2, we will outline our research design and identify the factors that have shaped the design. Next, we present in Section 3 the most important results from the WfM effectiveness study so far. Section 4 summarizes the lessons learned. Our conclusions and outlook are described in the final section.

2. Research design

2.1. Objective

The aim of the effectiveness study is to determine how the performance of the business processes is affected by the implementation of WfM technology. The four performance indicators selected to investigate for each involved business process are as follows:

- *Lead time*, i.e. the time between the arrival of a case and its completion (also known as cycle time, completion time, and turnaround time),
- *Service time*, i.e. the time spent by resources on the processing of a case,
- *Wait time*, i.e. the time a case is idle during its life cycle,
- *Utilization* of involved human resources, i.e. the ratio of activity versus their availability.

For each of these indicators, we report on the *average* values in this paper. We are aware of the importance of other measures, such as service levels and the degree of variance. These measures have been investigated, but for presentation purposes we focus on averages.

The specific indicators have been chosen on the basis of a literature review augmented with findings from several implementations in which we participated, which showed these to be the most popular quantitative metrics (Reijers, 2003). By introducing WfM technology, one may aim to decrease most of the average values of the given performance indicators. Because work is routed by an automated system, work reaches people faster and will not get lost. This decreases lead time and wait time. It will allow people to spend less time on coordination and on the transfer of work, which means a decrease of service time. When the supply of work and resources remain constant, work load and utilization will decrease as a result. Therefore, the hypothesis for this study is that the averages of all four performance indicators will decrease significantly as a result of the use of a WfM system.

Note that not under all circumstances it is desirable to achieve a low value for *each* performance indicator. Most notably, a low utilization seems favorable to accomplish a high level of flexibility, but some managers will rather aim for a high resource utilization to fully exploit the cost spent on labor.

2.2. Research steps

To determine the effects on process performance for each single process, at the very least, an initial measurement of the relevant parameters is required at two moments in time: (a) before the WfM implementation and (b) afterwards. In all, three major issues have further shaped the design of the research:

(1) The *validation* of the measurements: how can it be ensured that the collected data on a single process' performance is correct?
(2) The *prediction* of results: can we try to estimate the results of the WfM technology on a specific process before its actual implementation?
(3) The *comparison* of the measurements: how can a proper comparison between *ex ante* and *ex post* situations takes place?

The major steps in the research that address these issues are given in Fig. 1. In this figure, two axes can be distinguished. On the horizontal axis, we have the situation *before* the WfM technology implementation on the one hand and the situation *afterwards* on the other. On the vertical axis, we distinguish between the *real* data on a particular process on the one hand and the data that follow from a *simulation* of that process on the other. In the figure it is shown that there are six research steps, which take place in the order 0, 1a, 2a, 3, 2b, and 1b. We will explain these steps in detail and explain how they address the three issues of validation, prediction, and comparison. For now, it seems sufficient to say that the a-measurements use the initial circumstances, while b-measurements are based on the final circumstances.

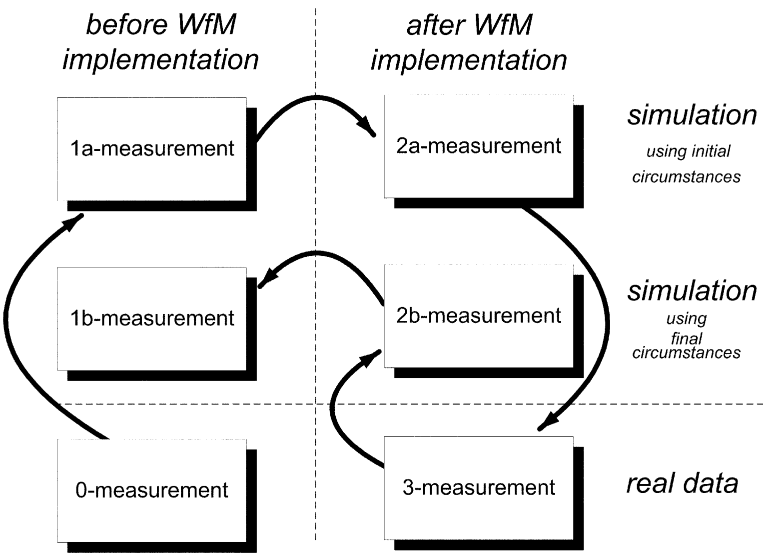

Fig. 1. Research steps.

The basis of the research design is formed by gathering *real* data on the process *before* and *after* the implementation of the WfM systems. We respectively refer to these measurements as the 0-measurement and the 3-measurement.

To address the issue of validation (1), a computer model is built of each business process subject to the study, both before and after the WFM implementation. We refer to the simulation of the model of the initial and final situation as respectively the 1a-measurement and the 2b-measurement. Both simulation models are as realistic as possible, including real data on the actual structure of the business process, the actual arrival of cases, the actual availability of resources, and the actual routing probabilities of cases flowing through the process, etc. Enactment of the simulation model delivers results on, for example, the lead times of the process and the resource utilization. These simulation results can be compared with the observations of the actual process. For example, the average lead time following from the execution of the simulation model can be compared with lead time averages as observed in practice.

Concordance of the real and simulated data gives support for the validity of the measurements, either of the initial situation or the situation after the WfM implementation. Significant differences between these outcomes may indicate that a part of the process is not understood or modeled correctly.

To enable prediction (2), we attempt to build a simulation model that reflects the situation after the implementation of the WfM system (the 2a-measurement). This model is based on the simulation model of the current process (used for the 1a-measurement) and captures both realistic and estimated data. On the one hand, it incorporates the aspects of the initial process that presumably will be the same after implementation. On the other hand, typical effects of WfM technology are incorporated in it. For example, transportation activities that exist in the current process are eliminated from the model, because WfM technology will take care of transportation (Jablonski & Bussler, 1996). Furthermore, planned initiatives of the organization to e.g. optimize the process structure or change the resource staffing are also incorporated in the model for the 2a-measurement. In this way, its estimate of the future overall effect is the most realistic. A comparison between the 1a- and 2a-measurement delivers insights in the expected benefits of the WfM technology.

The issue of comparison (3) is slightly more sophisticated. As we are primarily interested in the effect of the WfM technology, a straightforward comparison between the initial and final measurement (the 0- and 3-measurement) is perilous. After all, various context variables may have changed during WfM implementation that affect the final measurement. For example, if WfM technology is implemented while at the same time a staff reduction takes place, the performance established by the 0- and 3-measurement may be similar. It would not be proper in such a case to decide that WfM technology has not made any performance contribution. Similarly, the supply of work may have changed drastically.

To minimize the effect of these so-called *contextual* changes, we build a new simulation model that is used for the 1b-measurement. This model is derived from the 1a-model that reflects the initial situation. However, all contextual changes which happened during the WfM implementation are incorporated in the 1b-model as well. In the previous example of staff reduction, this would mean that the 1b-model includes e.g. the original process structure but a *reduced* number of staff compared to the original, initial situation. A comparison between the 1b and 2b-measurements will therefore be more meaningful, than a comparison between the 0- and 3-measurement.

In summary, the 1a- and 2b-measurements serve as validation for respectively the 0- and 3-measurement. A comparison between the 1a- and 2a-measurement gives as accurately as possible an estimation of the effects of WfM technology before implementation, while a comparison between the 1b- and 2b-measurement can be used to distinguish the nett effect of WfM technology from other contextual changes.

Note that an alternative and at first sight simpler research design would involve a large number of successive measurements during the entire implementation process and continuing for some time during steady state. In this way, changes in the performance indicators could be attributed more directly to exact events and simulations would become superfluous. However, the time and effort that is required to carry out a full measurement of process performance makes such a design infeasible.

2.3. Data gathering and analysis

Business processes contain a certain structure and they show a certain behavior. For this study, the most important categories of data to be determined for each business process are as follows:

- *Process*: tasks, milestones, business logic, routing probabilities
- *Resource*: types of resources, work assignment policies, number and availability of resources
- *Performance*: service times, lead times, arrival rate of new cases, work-in-progress, resource utilization.

For data gathering, the researchers used a multi-method approach (Yin, 1994), combining interviews, existing process descriptions, observations, management reports, self-registrations by people involved in the process, and automatically collected data by existing information systems. For each measurement, a careful consideration has been made for the most suitable mix of instruments.

An important difference between the 0- and 3-measurement with respect to data gathering concerns the availability of data. Wherever possible, the use of existing registrations on the processing of historic cases were favored over conducting new, manual registrations for reasons of representativity and ease of extraction. For the 0-measurement it was somehow inevitable that new data collection had to take place, for useful administration of this data within the organizations was often lacking (see Section 4.2). For the 3-measurement, the data gathered by the WfM system itself is an obvious rich and accessible source of this type of information.

Processes were modeled as Petri nets using the commercial tool Protos (Pallas Athena, 1997). The tool allows for efficient communication with end-users and the organization's management, thus simplifying knowledge extraction and validation. Protos models were automatically translated to simulation models, which could be executed and analyzed by the Petri-net based simulation tool ExSpect (Van Hee, Somers, & Voorhoeve, 1989). ExSpect provides a rich environment for simulation and analysis (e.g. confidence intervals, sensitivity analysis). For more information on the interplay between these tools, the reader is referred to e.g. Van der Aalst et al. (2000).

2.4. Progress

The workflow study started in September 2001 and is expected to continue until at least the beginning of 2007. Following from the nature of this type of research, the end data cannot be fixed

Table 1
Participants in the workflow study

Organization number	Organization description	Number of employees	Turnover/ budget (× millions €)	Focus of involved processes in study	Number of involved process in study	Cases per year (× 1000)
1.	Governmental agency	700	60 (b)	Debt collection	1	7000
2.	Health insurer	2300	5200 (t)	Policy maintenance	7	250
3.	Regional public works department	1000	250 (b)	Invoice processing	1	20
4.	Local municipality	300	210 (b)	Invoice processing	2	0.7
5.	Insurance intermediary	5000	29,000 (t)	Policy maintenance	3	2000
6.	Domiciliary care agency	1450	50 (t)	Human resource management	2	1.5

as the 3-measurements are dependent on the progress by the individual organizations on the implementation of their WfM technology. So far, six Dutch organizations have been actively involved in the study. A characterization of these organizations is given in Table 1. There is an equal balance of commercial and not-for-profit organizations involved in the study. Furthermore, both intermediate and large organizations are represented. Note that the column 'cases per year' shows the typical number of cases processed by the largest process under consideration in this study for that specific organization. The number of processes studied across all the organizations totals 16. This approaches the number of case studies involved in the study by Oba et al. (2000) and far exceeds the number of most other collected cases in research papers.

An organization could participate in the study when it had already selected a WfM system (but did not yet implement it). Normally, a WfM selection project involves considerable time and money. We used the criterion of WfM selection completion as an indicator of both the determination of the organization to complete the intended implementation and the remaining time to complete the implementation. The actual WfM systems involved in this study are three commercially available WfM systems (Staffware, COSA, and FLOWer) and one proprietary system (VenWfM).

For all listed organizations, the first half of the study (0-, 1a-, and 2a-measurement) has now been completed. For two of these, the second half of the study has been almost completed as well (1b-, 2b-, and 3- measurement). For two others, the WfM project has been stopped by the respective organizations. At the same time, three new candidate organizations have applied to participate in the study for which the study has yet to be initiated (not included in the list).

3. Results

A summary of the most important results is given in Table 2. The table gives the 1a- and 2a-measurements of the lead time and service time, as well as the 1a-measurement of the utilization.

Also shown are the expected gains from WfM systems for the lead time and service time, as can be derived from their 1a- and 2-measurements. Significant changes are accentuated.

For 15 out of 16 business processes (94%), the average lead time is expected to decrease significantly. The gains range from 25% to 83%, with an average of 48%. For process 16, the measured lead time reduction is not significant. Note that in this case the initial average lead time of one and a half day was already very small. For the business process involved, i.e. the processing of staff illness reports, this figure may indicate a natural lower bound on the lead time.

With respect to service time, for 12 out of 16 business processes (75%) a significant change is expected to take place. From these, 11 processes show an expected decrease of service time between 4% and 47%. However, in the situation of process 5 an *increase* of service time is expected to take place of 9%. On average, an expected decrease of service time of 22% is expected for the 12 processes which expressed significant changes.

It is interesting to take a closer look at process 5. It involves the handling of simple mutations of health insurance policies, such as caused by a change of address. It is the process with the lowest complexity and the lowest initial average service time value (3.51 min). Clearly, the overhead caused by the use of the WfM system—starting the system, registering work to be completed, etc.—can in this case not be compensated by less coordination work.

Table 2
Main results study

Org. nr. (see Table 1)	Proc. nr.	Lead time			Service time			Utilization
		1a-meas. (average value in days)	*2a-meas. (average value in days)*	*Reduction (%)*	*1a-meas. (average value in minutes)*	*2a-meas. (average value in minutes)*	*reduction (%)*	*1a-meas. (weighted average %)*
1.	1.	59.1	9.8	83**	13.45	7.14	47**	73
2.	2.	3.83	2.13	44**	16.01	9.01	44**	68
	3.	3.35	1.89	44**	4.16	4.01	4	65
	4.	3.40	1.96	42**	8.54	8.14	5*	65
	5.	3.76	1.72	54**	3.51	3.84	−9*	65
	6.	4.19	2.31	45**	9.25	8.90	4*	73
	7.	3.37	2.01	40**	10.75	8.19	24**	78
	8.	3.01	1.83	39**	5.4	3.89	28**	78
3.	9.	16.00	11.93	25**	17.66	17.39	2	4
4.	10.	6.50	1.81	72**	42.00	22.11	47**	3
	11.	13.08	6.82	48**	19.45	13.21	32**	36
5.	12.	6.17	4.56	26**	12.13	12.56	−4	60
	13.	5.17	2.34	55**	11.25	11.11	1	67
	14.	5.18	2.36	55**	12.06	11.03	9**	96
6.	15.	8.92	5.12	43**	24.19	20.97	13**	23
	16.	1.49	1.36	9	13.69	10.72	22**	71

*Significant with two-sided 90% confidence intervals.
**Significant with two-sided 99% confidence intervals.

Note that some categories of data are not shown in the table. In this phase of the study, they can still be derived from the presented data as follows:

- *The 0-measurements.*
  All average values of the 0-measurement are within the 99% confidence interval of the values of the 1a-measurement. In other words, the 1-measurements accurately reflect the situation at the 0-measurement.
- *The 2a-measurement on the utilization.*
  Utilization will change accordingly to the expected change of service time, because an equal supply of work and workforce is assumed after each WfM implementation.
- *The measurement on the wait time.*
  Because of the complete lack of concurrency, the wait time in each situation can be accurately determined by subtracting the service time from the lead time. The general relation between these entities is discussed in e.g. Reijers (2003).

In other words, the effects on utilization are equal to the effects on service time and the effects on wait time are equal to the effects on lead time. Note that in general these equivalences do not hold.

4. Lessons learned

The simulation study provided a detailed analysis of 16 administrative business processes. Although the main focus of the study was the effect of WfM technology on the four performance indicators described in Section 2.1, as a side-effect, we also gained insight in simulating business processes, gathering data, and structuring administrative business processes. In this section, we summarize the lessons learned.

4.1. Simulation

Although simulation has been used since the 1960s and today many mature simulation tools exist (Law & Kelton, 1991; Ross, 1991), the case study revealed quite some problems when it comes to mapping an administrative business process onto a simulation model. Simulation proved to be an effective validation means revealing many insights, but overall the application of the simulation tool turned out to be more difficult than expected. These difficulties were not caused by the use of a specific tool (the combination of Protos/ExSpect in this case), but by some more fundamental problems when mapping real-life business processes onto a simulation model. Therefore, we briefly discuss four of these generic problems.

The first problem concerns the fact that a simulation model typically focuses on a *single* process while the people involved distribute their time over *multiple* processes. This issue is also put forth in Sierhuis (2001) where it is stated that the multi-tasking behavior of human resources is often inadequately incorporated in simulation models. One way to address this issue is by not modeling a single process but the entire organization. The drawback is that the distribution of attention over processes is difficult to model, as it involves e.g. preferences, workload, and rewards.

Another way to address this problem is to simply focus on a single process and assume limited availability of resources (i.e., people). For example, if a person works on multiple processes, we can focus on a single process and abstract from other processes by assuming that the person is only available for a fraction of the time. The problem of the latter solution is that people typically do not spend a pre-specified fraction of their time on a specific process. It may be that during a period someone is working 50% on a process and during another for only 30%.

The second problem is related to the first problem in the sense that resources (i.e., people) are *not available* all the time, i.e., today many people work part-time. Even if people do not work part-time, they will have lunch breaks, holidays, and days off for training. Again, there are basically two approaches. The first one is to lower the number or resources, while the second is to make resources temporarily unavailable. Reducing the number of resources to reflect limited availability is not as easy as it seems. First of all, it is clear that there is a difference between 5 full-time resources and 10 half-time resources. Although the total number of full-time equivalents (FTE's) is five in both cases, the availability patterns are completely different. If there are 10 half-time resources, the average number of resources is five, but in principle there may be times that none of them is there while at other times 10 resources are available. This shows that 5 full-time resources cannot accurately model 10 half-time resources. Second, the average number of available FTE's does not have to be natural number, e.g., 7 half-time resources would be mapped onto 3.5 full-time resources. In such a situation, one still has to address the issue that resources are unavailable at times. Instead of reducing the number of resources, it is also possible to explicitly model the availability pattern of resources. Again this is not as simple as it appears to be. Clearly, there is a difference between a resource that can be used to up to 50% of the time and a resource that is simply not there for 50% of the time (e.g., only in the afternoons or only at certain days in the week). The first type of resources is clearly much more effective than the second. Note that in reality resources are in-between these two extremes, i.e., resources may be really unavailable at times but they may adapt this to current needs (e.g., someone postpones his "day off" when there is a backlog).

The third problem is the fact that people do not work at a constant speed. In a busy period people can process more cases. However, if people are offered too much work over a longer period of time, the performance tends to decrease. (Not to mention the effect on quality.) In psychological literature this effect is known as the Yerkes-Dodson law, which models the relationship between arousal, such as work pressure, and human performance as an inverted U-shaped curve, see e.g. Wickens (1992). So, unlike machines, human resources, have a performance that is rather elastic. This elasticity may depend on many circumstances ranging from workload and rewards to perceived stress and social structures. Obviously, it is difficult to model this in a simulation model in general. A simulation study in a manufacturing job-shop setting that incorporates a simple, variable relation between performance and work-load turned out to deliver results which were much more in line with empirical observations than previous simulations (Bertrand & Van Ooijen, 2002). It is an open question whether and how these findings may be generalized.

The fourth problem is that most business processes are confronted with varying workloads. There may be seasonal effects influencing the arrival process of new cases (e.g., an insurance intermediary and a travel agency will be more busy in December than in October). Process changes also cause unstable workloads. For example, a new law may take some time to implement, thus causing an accumulation of cases. Most simulation tools assume a steady-state process.

We think it is important to point out that the four problems are not so much of a technical nature. Using another simulation tool will not help. However, it would be interesting to think of new ways of modeling these phenomena. The first three problems are related to the availability of workers, the flexibility of this availability, and the effect of various workloads on this availability. A simple approach to tackle the first two problems would be to attach two attributes to each resource (in addition to standard attributes like role, etc.): (1) *availability* (a number between 0 and 1 denoting the percentage of time available) and (2) *timing elasticity* (a number between 0 and 1 denoting whether the distribution is fixed or depending on needs). Note that *performance elasticity* (e.g., the percentage of additional availability or the speed-up factor in case of too much work) is much more difficult to model since it is necessary to indicate under what circumstances the increased performance is available and for how long. The fourth problem is easier to address by using the real arrival process and feeding this to the simulation. Note that this approach does not allow for one long simulation run which is partitioned into subruns for statistical analysis (e.g., determining the confidence intervals for all kinds of metrics). This may be a problem for some tools since most simulation packages assume a steady-state process rather than a transient process.

4.2. Data gathering

For the 3-measurement it is typically possible to gather data from some transaction or event log. Most WfM systems record events such as the start and completion of tasks. Using this information it is possible to analyze the most relevant performance indicators. We have quite some experience in extracting such information from process logs (see e.g. Van der Aalst et al., 2003; Van der Aalst, Weijters, & Maruster, 2004). Recently, many vendors started offering tools for *Business Process Analysis* (BPA), cf. Sinur (2002). One of these tools is the ARIS Process Performance Monitor (ARIS PPM) which can analyze and visualize the performance of workflow processes based on some event log. Unfortunately, the event logs of some WfM products leave much to be desired. For example, Staffware (a WfM system used by several organizations involved in this project) does not log the start of a task, only its completion. Therefore, it is not possible to directly measure service times and waiting times. Using some heuristics it is possible to approximate these times but the only indicator that can be estimated directly is the lead time. Organization 3 (see Table 1) did use a custom-made WfM system. This system suffers from the same problem, i.e., only the transfer of work from one resource to another is monitored and therefore the actual start of a task is not measured. Again some heuristics and additional measurements where needed to estimate the actual service time and utilization. Despite these problems, it is relatively easy to gather data. In fact, the logs recorded by WfM systems offer many opportunities for detailed analysis. See Van der Aalst and Weijters (2004) for a discussion on process mining, i.e., extracting knowledge about resources, activities, cases, organizations, etc. from event logs.

Gathering data for the 0-measurement is often more problematic. Especially in the situation where paper documents are used rather than electronic documents, typically no or limited information on lead times, service times, and utilization is collected. The only way to address this problem is by recording the missing information by hand, i.e., people had to record the start time and end time of a task themselves (e.g., on some note attached to the document). Clearly, this is less accurate and it is labor-intensive. In fact, observation influences the process: it takes time to record start and end times and people tend to behave differently when being monitored. As a result, the quality of the data collected from 0-measurements was sometimes questionable. As much as possible, we used different data sources to cross-check the data. Also, we used the 1a-simulation model to explore the consistency between the various data. For example, self-reported utilization rates were checked against the utilization rates following from executing a simulation model which included self-reported service times and automatically gathered arrival rates.

Especially for the 0-measurements the available data-gathering period was often shorter than the usual lead time of a single case. Therefore, some heuristics were applied, e.g., instead of following single cases flow through the process to determine the service time spent on this case, more often for *each task* in the process its average service time was determined. This was done by counting a number of executions of the same task yet for different cases.

One of the lessons learned is to consider problems related to data gathering in an early phase of the project. Clearly, it is preferable to collect data automatically and build a process warehouse (Eder, Olivotto, & Gruber, 2002) that can be used for process mining purposes. One of the advantages of using a WfM system is that it is relatively easy to do this.

4.3. Process

An interesting side-effect of this study was that it gave the researchers the opportunity to examine the characteristics of business processes as executed, monitored, and designed in practice. We will highlight some of these insights here.

For each of the organizations, the performance criteria as distinguished in this study were mentioned as targeted by their own WfM implementation. This positively confirmed our ideas on the importance of these notions. Additional goals that were mentioned increased service quality, increased process flexibility, and a better integration of stand-alone applications.

One of the striking observations was that out of the 16 processes considered *none* of these processes incorporated concurrent behavior, i.e. parallel processing of single cases. Business processes turned out to be completely sequential structures. Their routing complexity was only determined by choice constructs and iterations. Even more remarkable is that for only one of these processes the process owners indicated that they considered to put more parallelism in the process once it would be supported by a WfM system. This contradicts with the idea that parallel processing is an obvious next step in improving the performance when adopting WfM technology (see e.g. Vander Aalst & Van Hee, 2002, p. 93). One of the explanations may be that processes allowing for parallelism have already been automated, i.e., the focus on organizations *starting* with a WfM system to replace "paper workflows" may explain this observation.

On a related note, the implementation of a WfM system did not prove to be a direct incentive to redesign the structure of the business process drastically. Without exception, each participating organization favored in the short term the situation to have a WfM system supporting the *current* process over a drastically *improved* version of the process. This may be counter-intuitive. One of the strengths of WfM technology's is that it enables the restructuring of the process structure (see e.g. Aversano, Canfora, De Lucia, & Galucci, 2002). Moreover, automating a "paper" process may not be the most effective way to achieve decreased lead and service times. On the other hand, this approach decreases the risk of failure by lowering the project's complexity. Therefore, the selected strategy may be favorable from a change management perspective.

4.4. Workflow paradox

As indicated in Section 2.4, two organizations did not implement a WfM system, i.e., the WfM project was stopped and we could not conduct a 3-measurement. The same phenomenon and in an even more intensified shape has been reported in Herrmann and Hoffmann (2003). This study describes several WfM projects that failed in the sense that the WfM was never actually introduced. Nevertheless, the authors note positive effects resulting from at least trying to introduce such a system. They state the following paradox: "It is sensible under certain circumstances to accept requests for workflow introduction and to commence such a project since this might be the most promising way leading to alternative solutions." (Herrmann & Hoffmann, 2003). We partly agree with this *workflow paradox*. However, we think that there are many organizations where a WfM system is introduced successfully. In the organizations where WfM projects are stopped, the decision is seldom based on technical problems but triggered by political or short-term financial arguments. One of the interesting phenomena we observed is that simply making a detailed model of the processes is very beneficial to the organization. Simulation forces organizations to reflect on their processes. For example, the insights from the simulation study conducted in Organization 2 resulted in a reduction of staff. This may trigger the conclusion that it is sufficient to conduct just a simulation study and forget about WfM systems. This conclusion is not valid for two reasons. First of all, many organizations benefit from the introduction of a WfM system. Second, without the real prospect of introducing a WfM system, the people involved will not take such a study very seriously. In other words, a WfM project is needed to "shake things up".

5. Conclusion and outlook

At this stage of the research, we have indications that WfM systems in general will positively affect the identified performance indicators averages. In the large majority of cases we investigated, service time and utilization are expected to decrease with 22%. For almost all cases, lead time and wait time are expected to decrease with more than twice that amount, namely 48%. Clearly, it needs to be seen whether these results will be accomplished in practice. On the basis of an almost completed 3-measurement for Organization 3, we are quite confident that the actual gains are indeed in the range of the predicted gains.

As a side-effect, this empirical study has proved to be a valuable source of information on actual business process properties and their execution. Also, simulation proved to be a good way of validating the initial measurements, but a number of challenges had to be faced. We have addressed directions to counter some of the identified problems.

Unfortunately, we have seen two organizations putting their WfM implementations on hold, perhaps definitively. We are still attracting new organizations to get involved in the study to generate support for general conclusions.

Currently we are carrying out the first two 3-measurements. As part of this work, we are developing a general tool to derive from the event logs of WfM systems performance information. We are studying industrial solutions in this field, such as the ARIS PPM and XML data formats.

We sometimes have the impression that research in the WfM area seems to focus on the aspects of the technology which are still less-than-perfect, e.g. the limited capabilities to cope with exceptions and the problems associated with straight-jacketing inherently fluid processes in executable form. We do not deny that these are important areas for further research and improvement. Yet, we hope that the results from this first part of our study—as well as from what is to follow—will contribute to a more balanced and better supported evaluation of the industrial value of WfM technology. After all, if our predictions will hold, WfM technology is effective in improving process performance, even while there may still be things to be desired about these systems.

Acknowledgements

The following people from Eindhoven University of Technology and Deloitte have helped in setting up the research, gathering the data and its analysis: Rochelle Degens, Dick van Gorkum, Kees van Hee, Harm Hoebergen, Charo Narváez, Oscar Ommert, René Theunissen, Alexander in ‘t Veld and Eric Verbeek.

References

Van der Aalst, W. M. P., De Crom, P. J. N., Goverde, R. R. H. M. J., VanHee, K. M., Hofman, W. J., Reijers, H. A., et al. (2000). ExSpect 6.4: An executable specification tool for hierarchical colored Petri nets. In M. Nielsen, & D. Simpson (Eds.), *Application and theory of Petri nets 2000. Lecture notes in computer science*, Vol. 1825 (pp. 455–464). Berlin: Springer.

Van der Aalst, W. M. P., Van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A. J. M. M. (2003). Workflow mining: A survey of issues and approaches. *Data and Knowledge Engineering*, *47*(2), 237–267.

Van der Aalst, W. M. P., & Van Hee, K. M. (2002). *Workflow management: Models, methods, and systems*. Cambridge: MIT Press.

Van der Aalst, W. M. P., & Weijters, A. J. M. M. (2004). Process mining: A research agenda. *Computers in Industry*, *53*(3), 231–244.

Van der Aalst, W. M. P., Weijters, A. J. M. M., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1128–1142.

Aversano, L., Canfora, G., De Lucia, A., & Galucci, P. (2002). Business process reengineering and workflow automation: A technology transfer experience. *The Journal of Systems and Software*, *63*(1), 29–44.

Bertrand, J. W. M., & Van Ooijen, H. P. G. (2002). Workload based order release and productivity: A missing link. *Production and Planning control*, *13*(7), 665–678.

Dourish, P. (2001). Process descriptions as organizational accounting devices: The dual use of workflow technologies. In C. A. Ellis, & I. Zigurs (Eds.), *Proceedings of the ACM 2001 international conference on supporting group work* (pp. 52–60). New York: ACM Press.

Eder, J., Olivotto, G. E., & Gruber, W. (2002). A data warehouse for workflow logs. In Y. Han, S. Tai, & D. Wikarski (Eds.), *Proceedings of the international conference on engineering and deployment of cooperative information systems* (pp. 1–15). Berlin: Springer.

Ellis, C. A. (1979). Information control nets: A mathematical model of office information flow. In P. F. Roth, & G. J. Nutt (Eds.), *Proceedings of the ACM conference on simulation, measurement and modeling of computer systems* (pp. 225–240). New York: ACM Press.

Goebl, W., Messner, K. J., & Swarzer, B. (2001). Experiences in introducing workflow management in a large insurance group. In J. R. Sprague, & H. Ralph (Eds.), *Proceedings of the 34th Hawaii international conference on system sciences* (p. 9037). Washington DC: IEEE.

Van Hee, K. M., Somers, L. J., & Voorhoeve, M. (1989). Executable specifications for distributed information systems. In E. D. Falkenberg, & P. Lindgreen (Eds.), *Proceedings of the IFIP TC 8/WG 8.1 working conference on information system concepts: An in-depth analysis* (pp. 139–156). Amsterdam: Elsevier Science.

Herrmann, T., & Hoffmann, M. (2003). The metamorphoses of workflow projects in their early stages. Submitted to CSCW—The Journal of Collaborative Computing, 37pp. Available at: http://iugsun.cs.uni-dortmund.de:2048/pub/bscw.cgi/0/208299/30310/30310.pdf.

Jablonski, S., & Bussler, C. (1996). *Workflow management: Modeling concepts, architecture, and implementation*. London: International Thomson Computer Press.

Kueng, P. (2000). The effects of workflow systems on organizations: A qualitative study. In W. M. P. Van der Aalst, J. Desel, & A. Oberweis (Eds.), *Business process management: Models, techniques, and empirical studies* (pp. 301–316). Berlin: Springer.

Law, A., & Kelton, D. (1991). *Simulation modeling and analysis*. New York: McGraw-Hill.

Lawrence, P. (Ed.). (1997). *Workflow handbook 1997*. New York: Wiley.

Oba, M., Onada, S., Komoda, N. (2000). Evaluating the quantitative effects of workflow systems based on real cases. In *Proceedings of the 33rd Hawaii international conference on system sciences*, Vol. 6 (p. 6031). Washington DC: IEEE.

Pallas Athena. (1997). *PROTOS user manual*. Plasmolen: Pallas Athena BV.

Parkes, A. (2002). Critical success factors in workflow implementation. In T. Terrano, & M. D. Myers (Eds.), *Proceedings of the sixth Pacific Asia conference on information systems* (pp. 363–380). Tokyo: Jasmin.

Prinz, W., & Kolvenbach, S. (1996). Support for workflows in a ministerial environment. In M. S. Ackerman (Ed.), *Proceedings of the ACM conference on computer supported cooperative work* (pp. 199–208). New York: ACM Press.

Reijers, H. A. (2003). *Design and control of workflow processes: Business process management for the service industry. Lecture notes in computer science*, Vol. 2617. Berlin: Springer.

Ross, S. M. (1991). *A course in simulation*. New York: Macmillan.

Sierhuis, M. (2001). *Brahms: A multiagent modeling and simulation language for work system analysis and design*. Ph.D. thesis, Department of Social Sciences and Informatics, University of Amsterdam, Amsterdam.

Sinur, J. (2002). *BPA market catches another major updraft*. Gartner's application development and maintenance research note M-16-8153. Gartner, 12 June 2002.

Staffware (2000). Benefits, progress and competitive edge—the customer's experience. Staffware, Maidenhead, September 2000.

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: HarperCollins.

Wintergreen. (2003). *Business process management (BPM) market opportunities, strategies, and forecasts, 2003 to 2008*. Lexington: WinterGreen Research.

Yin, R. K. (1994). *Case study research, design and methods* (2nd ed.). Newbury Park: Sage Publications.

Hajo Reijers is an assistant professor in Information Systems at the Faculty of Technology Management of Eindhoven University of Technology. He is also a staff member of the Process Management Research Center, Babson College. Previously, he has worked for several management consultancy firms, most recently as a manager with Deloitte. His main interests are business process redesign, information systems, workflow management systems, Petri nets, and simulation.

Wil van der Aalst is a full professor of Information Systems and head of the Information Systems department of the Faculty of Technology Management at Eindhoven University of Technology. Currently he is also an adjunct professor at Queensland University of Technology (QUT) working within the Centre for Information Technology Innovation (CITI). His research interests include information systems, simulation, process mining, Petri nets, process models, workflow management systems, verification techniques, enterprise resource planning systems, computer supported cooperative work, and interorganizational business processes.