On the performance of workflow processes with distributed actors : does place matter?

Citation for published version (APA):

Reijers, H. A., Song, M. S., & Jeong, B. (2007). On the performance of workflow processes with distributed actors : does place matter? In G. Alonso, P. Dadam, & M. Rosemann (Eds.), *Proceedings of the 5th International Conference on Business Process Management (BPM 2007) 24-28 September 2007, Brisbane, Australia* (pp. 32-47). (Lecture Notes in Computer Science; Vol. 4714). Springer. https://doi.org/10.1007/978-3-540-75183-0_3

DOI:

[10.1007/978-3-540-75183-0_3](https://doi.org/10.1007/978-3-540-75183-0_3)

Document status and date:

Published: 01/01/2007

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

On the Performance of Workflow Processes with Distributed Actors: Does Place Matter?

Hajo A. Reijers¹, Minseok Song¹, and Byungduk Jeong^{1,2}

¹ Department of Technology Management, Eindhoven University of Technology,
P.O.Box 513, NL-5600 MB, Eindhoven, The Netherlands
{h.a.reijers,m.s.song}@tue.nl

² Department of Industrial Engineering, Korea Advanced Institute of Science and
Technology, South Korea
tauruseo@kaist.ac.kr

Abstract. Current workflow technology offers rich features to manage and enact business processes. In principle, the technology enables actors to cooperate in the execution of business processes regardless of their geographical location. Furthermore, the technology is considered as an efficient means to reduce processing times. In this paper, we evaluate the effects on the performance of a workflow process in an organizational setting where actors are geographically distributed. The studied process is exceptional, because equivalent tasks can be performed at different locations. We have analyzed a large workflow process log with state-of-the-art mining tools associated with the ProM framework. Our analysis leads to the conclusion that there is a positive effect on process performance when workflow actors are geographically close.

Keywords: Workflow management, performance evaluation, process mining, case study.

1 Introduction

Since the mid 1990s, Workflow Management Systems (WfMSs) have received wide attention as a research subject in the IS community. Recent interest for Process-aware Information Systems [11] breathes new life into WfMS's fundamental concept, the distribution of work to people and systems on the basis of a pre-defined process model [1]. The industrial success of WfMS's can be clearly seen in, for example, the Netherlands and South-Korea, where every bank, insurance company, ministry, and most municipalities have adopted this technology.

Nonetheless, little is known about the extent to which workflow technology helps organizations to execute their business processes more efficiently and effectively. Market analysts and software vendors boast success stories, but they hardly play an impartial role in this discussion. To fill this white space, a research project was initiated in 2001 by Eindhoven University of Technology and Deloitte Consultancy. The purpose of the project was to involve as large a number of Dutch organizations as possible to closely monitor their experiences with implementing and using WfMSs to support their business processes over a period of time. For an overview of the preliminary results, the reader is referred

to [22]. This project, in which 10 organizations are involved and over 20 business processes, is currently nearing its completion.

One of the organizations that participated in the mentioned project is a Dutch municipality, which started the implementation of the WfMS Staffware for their invoice handling process in 2003. In the second half of 2004, they went “live”. The research project we mentioned gave us access to the WfMS’s process log which contained the registered events for 2005’s production, covering over 12,000 completed cases (invoices) in total. The interesting thing is that there are tasks which can be executed at 10 different locations all across the city. After all, an invoice in this setting can pertain to almost anything (e.g., pencils, pc’s, or furniture) and it must be checked by the responsible civil servant who issued it. Such a civil servant may be working at the city’s fire brigade, swimming pool, theater, or any of the other locations. So, this implementation site provided a rare opportunity to evaluate whether it matters for process performance when actors are geographically distributed.

In this paper, we focus on processing time and transfer time as performance measures to investigate whether a relation exists between geographical location and process performance. As far as we know, this has not been investigated before in the setting of an actual WfMS implementation on the basis of real data. But insight into this relation - if it exists - may be valuable to manage and influence the performance of future workflow implementations.

In the next section we will start with an overview of related work. In Section 3 we will describe our research design, in particular the hypotheses we set out to investigate. Then, we will describe the case study in more detail and report our analysis and findings in in Section 4. The paper ends with a discussion and some concluding remarks.

2 Related Work

2.1 Workflow and Geography

By having a WfMS in place for the logistic management of a business process, such processes can theoretically be executed faster and more efficiently [19]. WfMS vendors claim that these advantages materialize in practice. In academic papers, various single case studies of workflow implementations are described and a small number of studies that involve multiple implementations [15,18,21]. Most of the studies that explicitly consider performance established a positive effect of workflow technology, in particular in [21]. However, none of these studies examined whether the geographical distribution of actors played any part in such performance improvement. (Note that the *architectural* issues that relate to distributed workflow processing have been widely studied, e.g. in [12,13,20]).

It was established in the seminal work by Thomas Allen at MIT [4] that geographical distances between actors may matter. In the late 1970s, Allen undertook a project to determine how the distance between engineers’ offices coincided with the level of regular technical communication between them. The results of

that research, now known as the *Allen Curve*, revealed that when there is more distance between people they will communicate less frequently.

However, it is believed that due to the massive utilization of information and communication technologies (ICTs) the precise physical location of individual participants will become irrelevant to their interactions [7,8]. ICTs are a key enabler for the emergence and sustained popularity of so-called virtual teams, i.e. groups of geographically and organizationally dispersed coworkers that are assembled using a combination of telecommunications and information technologies to accomplish an organizational task [25]. WfMSs too enable the fast communication and collaboration between geographically dispersed users and can therefore be expected to contribute to improved interaction between them [6,23,24]. In particular, in [1] it is stated that “The introduction of a WfMS lowers the physical barriers between the various sections of an organisation”. It continues that a WfMS can, for example, be used to more evenly distribute work among geographically scattered resources. Therefore, we may assume from existing literature that it is less relevant where people reside physically for the performance of a process that is managed by a WfMS.

2.2 Process Mining

In this paper, we use process mining techniques to analyze business process execution results. Process mining allows the discovery of knowledge based on a process log [3]. The process log, which is provided by most process aware information systems, records the execution of tasks in some business processes. Process mining can deal with several perspectives, such as the process perspective, organizational perspective, performance perspective, etc. To support process mining, several tools have been developed [14,16,2,10]. The ProM framework has been developed to support various process mining algorithms such as these. It was designed to easily add new algorithms and techniques into it by means of plug-ins[10]. A plug-in is basically the implementation of an algorithm that is of use in the process mining area.

Figure 1 shows an overview of the ProM framework. ProM reads log files in the XML format through the Log filter component. This component can handle large data sets and sort the events within a case according to their time stamps. Through the Import plug-ins a wide variety of models can be loaded, ranging from Petri nets to logical formulae. The Mining plug-ins perform the actual process mining. The Analysis plug-ins take a mining result and perform an analysis. The ProM framework provides several analysis techniques such as Petri-net analysis, social network analysis, performance analysis, etc. The Conversion plug-ins can convert a mining result into another format, e.g., from an EPC into a Petri net.

3 Research Design

This section explains our research questions and describes the research method with which they are addressed. The objective of this study is to determine how

Fig. 1. Overview of the ProM framework [10]

the performance of a business processes is affected by the use of a WfMS in a geographically distributed setting. Before explaining our research questions, the two process performance indicators selected to investigate are defined as follows:

- *Processing time*: the time between the start of a task and its completion,
- *Transfer time*: the time between the completion of a task and the start of a subsequently executed latter task

Applying a WfMS could result in a reduction of processing times, because it delivers the right work to the right person at the right time. When a WfMS takes care of assigning work to actors, it is perhaps less relevant where these actors are located geographically. When companies introduce WfMSs, they normally perform business process re-engineering projects. During the projects, they carry out as-is analyses and try to remove the geographical influences in the execution of business processes by standardizing the tasks in the business processes. After that, they design new business processes and implement them with WfMSs. In WfMSs, when a task is completed, the task is immediately assigned to a proper actor (the worklist of the actor) in spite of his/her geographical location. Next, it is handled by the actor. Thus, it seems that the introduction of workflow technology takes away any geographical influences. To evaluate this argument, we established a research procedure as shown in Figure 2. Under this procedure, we examine whether processing and transfer times are affected by workflow technology in terms of the geographical location of its involved actors.

The first step is generating our hypotheses. Our research questions led to the formulation of two hypotheses:

- *Hypothesis 1*: The processing time of equivalent tasks is equally distributed, despite the geographical locations in which the tasks are performed.

Fig. 2. The research procedure

- *Hypothesis 2*: The transfer time between tasks within the same geographical location is equally distributed as the transfer time between tasks across geographical locations.

Hypothesis 1 deals with the first research question. For the hypothesis, we considered tasks that can be performed in several geographical locations. We calculated processing times of tasks within each location and compared them. *Hypothesis 2* addresses the second question. In this case, we took into account the pairs of tasks that can be successively executed in the same geographical location or across different geographical locations. Note that geographical separation may also lead to time differences between locations. However, in this paper, this is not the case.

After generating the hypotheses, we gathered process logs from the involved organization, which operates Staffware as its WfMS. Since the process logs gathered were stored in a proprietary format, we had to preprocess the process logs. They were converted into a standard MXML format [10]. After the conversion, we analyzed them with the ProM framework and its associated tools. We removed irrelevant tasks and calculated relevant processing and transfer times. After that, we performed various statistical tests to examine our hypotheses. Since the ProM framework does not support statistical analysis, we generated the data for statistical analysis from ProM and used Statgraphics Centurion XV for the tests. The analysis results from ProM and the statistical test results were reported to the organization that provided the logs. Finally, we gathered their feedback.

4 Case Study

4.1 Context

The context of our case study is the Urban Management Service of a municipality of 90,000 citizens, situated in the northern part of the Netherlands. The municipality is one of the organizations that is involved in our longitudinal study into the effectiveness of workflow management technology [22]. In 2000, the board of the municipality decided to implement a WfMS throughout the organization,

which encompasses some 300 people. Mainly because of restricted budgets and some technical setbacks, it lasted until 2004 before the first two business processes were supported with this technology. One of these two processes involves the handling of invoices, which is the focus of our analysis.

On a yearly basis, the municipality deals with some 20,000 invoices that pertain to everything that the municipality purchases. The overall process consists of 26 tasks and may involve almost every employee of the Urban Management Service. After all, an important check is whether the invoice is ‘legitimate’ in the sense that it corresponds with an authorized purchase by some employee, to be checked by that employee himself/herself. The general procedure is that an invoice is scanned and subsequently sent by the WfMS to the central financial department. A clerk registers the invoice after which it is sent to the proper local financial office. These local financial offices are distributed over all the geographical locations of the municipality (e.g. the mayor’s office, the city’s swimming pool, the fire brigade, etc.). Depending on the kind of invoice, there are various checks that need to take place: the person responsible for the budget that is used for the purchase must approve (the budget keeper); the fit between the purchase with the supplier’s contract (if any) must be established; various managers may be required to authorize the invoice depending on the amount of the money involved; etc. Eventually, a purchase may be paid by the central financial office.

4.2 Analysis Procedure

In the case study under consideration, a process log is automatically generated by the WfMS executing the invoice handling process. A process log consists of several *instances* or *cases*, each of which may comprise several audit trail entries. An *audit trail entry* corresponds to an atomic *event* such as schedule, start, or completion of a task. Each audit trail entry records task name, event type, actor and time stamp. Figure 3 shows the example of translated process logs. In the figure, the names of actors are replaced by artificial IDs to ensure confidentiality and privacy. The process log starts with the *WorkflowLog* element that contains *Source*, and *Process* elements. The *Source* element refers to the information about the software or the system that was used to record the log. In our case study, the log comes from a “Staffware” system. The *Process* element represents the process to which the process log belongs. *ProcessInstance* elements correspond to cases. The *AuditTrailEntry* element represents a log line. It contains *WorkflowModelElement*, *EventType*, *Timestamp*, and *Originator* elements. The *WorkflowModelElement* refers to the activity the event corresponds to. The *EventType* specifies the type of the event, e.g., *schedule* (i.e., a task becomes enabled for a specific instance), *start* (the beginning of a task instance), and *complete* (the completion of a task instance), etc. The *Timestamp* refers to the time when the event occurred and the *Originator* corresponds to the originator who initiates the event.

The process log we analyzed covered slightly more than 12,000 instances (completely handled invoices), as processed by the municipality in the first half of 2005. This pertained to a huge amount of data: It has more than 200,000 events

```

<?xml version="1.0" encoding="UTF-8"?>
<WorkflowLog xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="http://www.is.tn.tue.nl/research/processmining/
    WorkflowLog.xsd">
  <Source program="Staffware"/>
  <Process id="Facturen" description="none">
    <ProcessInstance id="4-21334" description="none">
      <AuditTrailEntry>
        <WorkflowModelElement>PREREG</WorkflowModelElement>
        <EventType>schedule</EventType>
        <Originator>actor1</Originator>
        <Timestamp>2005-07-13T13:55:00+01:00</Timestamp>
      </AuditTrailEntry>
      <AuditTrailEntry>
        <WorkflowModelElement>CODFCTBF</WorkflowModelElement>
        <EventType>start</EventType>
        <Originator>actor2</Originator>
        <Timestamp>2005-07-13T13:58:00+01:00</Timestamp>
      </AuditTrailEntry>
      <AuditTrailEntry>
        <WorkflowModelElement>CODFCTBF</WorkflowModelElement>
        <EventType>complete</EventType>
        <Originator>actor2</Originator>
        <Timestamp>2005-07-13T14:01:00+01:00</Timestamp>
      </AuditTrailEntry>
    ...
  </ProcessInstance>
</Process>
</WorkflowLog>

```

Fig. 3. An example log

and about 350 actors are involved in the process. We investigated the whole process log and decided to focus our attention to two specific elements. First of all, we decided to analyze the processing times of five specific *tasks*, being the most important checks as prioritized by the financial management. The five tasks are CODFCTBF, CONTRUIF, ROUTEFEZ, CONTRCOD, and FBCONCOD. Note that the five tasks can be performed in several geographical locations. We left out administrative tasks like scanning, keying in data, categorizing, archiving, etc. Secondly, we considered four *pairs* of tasks where we could establish that at times they were subsequently performed within the same geographical unit and at other times across different units. They are ROUTEFEZ-CODFCTBF, CODFCTBF-CONTRUIF, CODFCTBF-CONTRCOD, and CONTRCOD-BEOORDSR.

We calculated the processing time of each task and the transfer time of each pair. Before the actual mining starts, the process data was filtered to focus on a specific task or pair. The ProM framework provides several filters that enable the removal of irrelevant information from process logs. For example, the *event log filter* is used to extract the events in which we are interested. If we apply the filter to the log in Figure 3 and filter out the ‘PREREG’ activity, we obtain the log in Figure 4 where the ‘PREREG’ activity is removed. Besides the *event log filter*, we also applied several filters to preprocess the log and improve its analyzability.

After applying the filters, we used the performance sequence diagram analysis plug-in. This plug-in makes a sequence diagram from process logs and shows performance measures such as average throughput time, transfer time, time spent in a task, etc. A sequence diagram has vertical and horizontal dimensions. The


```

<?xml version="1.0" encoding="UTF-8"?>
<WorkflowLog xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="http://www.is.tm.tue.nl/research/processmining/
    WorkflowLog.xsd">
  <Source program="Staffware"/>
  <Process id="Facturen" description="none">
    <ProcessInstance id="4-21334" description="none">
      <AuditTrailEntry>
        <WorkflowModelElement>CODFCTBF</WorkflowModelElement>
        <EventType>start</EventType>
        <Originator>actor2</Originator>
        <Timestamp>2005-07-13T13:58:00+01:00</Timestamp>
      </AuditTrailEntry>
      <AuditTrailEntry>
        <WorkflowModelElement>CODFCTBF</WorkflowModelElement>
        <EventType>complete</EventType>
        <Originator>actor2</Originator>
        <Timestamp>2005-07-13T14:01:00+01:00</Timestamp>
      </AuditTrailEntry>
    </ProcessInstance>
  </Process>
</WorkflowLog>

```

Fig. 4. The filtered log

Fig. 5. ProM screenshot showing sequence diagram (pattern view)

vertical dimension is a time dimension and the horizontal dimension shows classifier roles that represent geographical locations.

Figure 5 shows the sequence diagram of transfer time for the ROUTEFEZ-CODFCTBF pair. In the figure, there are two kinds of patterns, such as boxes and arrows. When a transfer happens within a geographical location, it is represented as a box. If it happens across different geographical locations, an arrow between them is drawn. As shown in the figure, the transfer times also vary according to the various geographical locations.

With the ProM framework, we can calculate performance indicators (i.e. average time, minimum time, maximum time, and standard deviation) of both

processing and transfer times. We exported the analysis result of the ProM framework to Statgraphics and performed statistical analysis.

4.3 Analysis and Findings

Processing time. For our analysis, we determined the average processing times of all five tasks under consideration. The results for the CODFCTBF task, which covers the largest number of different geographical locations, is shown in Figure 6. The task involves the check on the legitimacy of the invoice by the responsible budget keeper.

Fig. 6. Average processing time of the CODFCTBF task

The figure shows that the average processing times for the CODFCTBF task differ across the various geographical locations. These averages range between the extremes of approximately 10 hours and 53 hours. Although the CONTRUIF, ROUTEFEZ, CONTRCOD and FBCONCOD tasks involve fewer geographical locations – respectively only 3, 7, 6 and 2 – the variation is similar to the CODFCTBF task.

Using the Kolmogorov-Smirnov test, we could reject with a 95% reliability that processing times of any task were normally distributed. This violates the assumptions for most standard parametric tests to determine statistical differences (e.g. ANOVA), which explains our use of the distribution-free Kruskal-Wallis test that compares medians. For all tasks under consideration, this test leads with a 95% confidence to the outcome that there is a *significant difference between the processing times across various locations*. Because of the existence of outliers, we also applied Mood’s median test, which is less powerful but more robust in this respect: It leads to the same result.

To illustrate the relative difference within the processing times for a single task, we present a Box-and-Whisker plot (also known as *boxplot*) for the CODFCTBF task in Figure 7. In the plot, the medians are shown as notches between the lower and upper quartiles. The plot suggests differences between, for example, the medians of locations 1 and 3, locations 2 and 4, etc.

To investigate whether the found differences in processing times across the geographical locations persist over time or are rather of a transient nature, we

Fig. 7. Box-and-Whisker plot for the CODFCTBF task

split up the overall log in 6 chronologically subsequent smaller logs of equal size and analyzed these as well. As additional analyses confirmed the non-normality of the processing times within all sublogs, we again used the Kruskal-Wallis test. The result is shown in Table 1. Note that the columns from the second to the seventh represent each sublog.

Table 1. The Kruskal-Wallis test result (processing time), significant differences at a 95% confidence interval indicated with ‘*’

task	1	2	3	4	5	6
CODFCTBF	*	*	*	*	*	*
CONTRUIF	–	*	*	–	–	–
ROUTEFEZ	*	*	*	*	*	*
CONTRCOD	*	*	*	*	*	*
FBCONCOD	*	*	*	*	*	*

What can be seen is that for all but the CONTRUIF task the processing times across the locations vary significantly at a 95% confidence level for *all* its sublogs. (For the CONTRUIF task, this difference is only significant for the 2nd and 3rd sublog and is therefore considered as being of a transient nature.) So, we reject our first hypothesis. Processing times tend to differ significantly across the geographical locations where they are performed and do so for successive periods of time.

Transfer time. For analyzing the transfer time, we focus on the four pairs of tasks we mentioned earlier. The transfer points were selected because on the basis of a content analysis of the tasks it can be imagined that the involved tasks either take place entirely within the same geographical location or that each task is carried out in a different location. In the first case, we speak of

an *intra transfer*, as the work is transferred between executors within the same location; in the second case, an *inter transfer*, as the executors are at different locations.

An analysis of the process log indicated that only for two of the pairs where intra and inter transfers take place there is sufficient data to compare these transfers in a meaningful way. For the other two, there are at most 50 observations of inter transfers versus thousands of observed cases for inter transfers. Therefore, we focus on the following two pairs:

1. *from ROUTEFEZ to CODFCTBF*: the initial check by a local financial clerk whether an invoice is intended for the sector that the clerk is attached to, and if so, the subsequent check on the legitimacy of the invoice by a budget keeper;
2. *from CODFCTBF to CONTRCOD*: the legitimacy check by a budget keeper followed by the check of a local financial clerk whether the control code as filled out by the budget keeper is correct;

For these pairs, there are respectively 2125 and 1764 inter transfers and approximately three times as many intra transfers within each category.

Application of the Kolmogorov-Smirnov test points out that with a 95% confidence the idea can be rejected that transfer times for either pair are normally distributed. This makes a test that focuses on the comparison of medians of the transfer times more suitable. Figure 8 shows that for both transfer types, the median of the inter transfer time exceeds that of the intra transfer time, although this difference is larger in the case of transfers from ROUTEFEZ to CODFCTBF.

Fig. 8. Median transfer times at transfer points

Similar as for the analysis of the processing times, the Kruskal-Wallis test was selected to test the equality of medians between intra and inter transfers. In the presence of outliers, Mood's median test was applied as a more robust yet less powerful, additional test. For *both* transfer types, the Kruskal-Wallis test shows significant differences between intra and inter transfers at a 95% confidence interval. At the same confidence level, Mood's median test only shows a significant difference for the transfer of work from ROUTEFEZ to CODFCTBF.

For both transfer types, box-and-whisker plots that show the area between the lower quartile and upper quartile of the data values with the median in the middle, are given in Figures 9. Small markers (plus signs) indicate the means for intra and inter transfer times.

Fig. 9. Detail Box-and-Whisker plots for the transfer of work

So, both statistical tests point at a *significant difference* between the intra and inter transfer times for the transfer of work from ROUTEFEZ to CODFCTBF, where inter transfers clearly exceed intra transfer times. The approximate confidence intervals for the medians, indicated by the notches in the quartile bodies in the Box-and-Whisker plot, confirm this result as they are wide apart and do not overlap. The difference is not so apparent for work being transferred from CODFCTBF to CONTRCOD.

Finally, to determine whether the differences between the intra and inter transfer times persist over time, the complete log is split up in 6 subsequent smaller logs of equal size. The procedure is similar as in the case of the analysis of processing times, as described earlier in this section. As Kolmogorov-Smirnov tests confirmed the non-normality of the transfer times times within all sublogs, we again used the Kruskal-Wallis test. The result is shown in Table 2.

Table 2. the Kruskal-Wallis test result (transfer time), significant differences at a 95% confidence interval indicated with ‘*’

task pair	1	2	3	4	5	6
From ROUTEFEZ to CODFCTBF	*	*	*	*	*	*
From CODFCTBF to CONTRCOD	-	*	*	-	*	*

For the pair ROUTEFEZ-CODFCTBF the significant difference between intra and inter transfers is present in *all* sublogs. Therefore, we reject our second hypothesis. After all, for this pair at least we see that intra and inter transfer times vary significantly over a long period of time.

4.4 Discussion

Summary of findings. We rejected our hypotheses that suppose that workflow technology takes away geographical barriers (Hypotheses 1 and 2). With respect to the first part, our analysis shows significant differences between processing times of equivalent tasks across different geographical locations; similarly significant differences between intra and inter transfer times are found.

Evaluation. We gathered feedback on the found results from a team of the involved municipality, which included the financial manager, functional administrator of the workflow system, a systems integrator responsible for technical modifications, and a budget-keeper/executor. We had a one-and-a-half hour meeting with them in the city town-hall, where we presented and discussed the results, followed-up by several e-mail contacts and phone conversations.

No satisfactory explanation could be found for the surprising differences in processing times (Hypothesis 1), as the team members once more confirmed that the tasks are strictly equivalent across the various locations. Differences in local skills and perhaps informal norms may contribute to the difference. This is in line with research in the tradition of “social ecology” [5]. It positions that different social settings, such as offices and meeting rooms, are associated with different behavioral norms, mental schemas, and even scripts that sharply affect the way people act and the expectations they have of others.

After considerable deliberation, a possible explanation was found for the difference in transfer times (Hypothesis 2). Within the municipality, local financial clerks are provided with reports on “open” invoices. These can be used to urge budget keepers to check the invoices that are with them for some time. The team from the municipality suspects that this encouragement is done more frequently and more persuasively in settings where the clerks and budget keepers are in the same location, which may well explain the distinctive difference between transfer times from ROUTEFZ and CODFCTBF.

But even if the encounters between financial clerks and budget keepers are not planned, the effect of spontaneous communication between them should not be underestimated. It seems logical that spontaneous encounters will take place more frequently when people reside in the same building. With spontaneous casual communication, people can learn, informally, how one another's work is going, anticipate each other's strengths and failings, monitor group progress, coordinate their actions, do favors for one another, and come to the rescue at the last minute when things go wrong [9]. But physical separation drastically reduces the likelihood of voluntary work collaboration [17].

Limitations. Clearly, this study is carried out within the setting of a single organization, so the usual limitations apply with respect to generalizing its results.

A more specific concern could be raised on the validity of reasoning over process performance, as we strongly focused on the analysis of an automatically generated process log. Obviously, process logs are by no means a full representation of what is going on in an organization. However, for the reported case it seems likely that the recorded events follow actual work execution quite closely,

as confirmed by the team of the municipality. In another part of the larger research project we are involved in (see [22]), we have seen an implementation where people worked around the workflow system on a wide scale, e.g. using the workflow system in batch mode to check out work that was completed manually much earlier. In such a case, it would be much more dubious to draw conclusions of the kind we did. The patterns in the process logs that hinted at such anomalous behavior, i.e. (1) extremely short processing times and (2) many “bursts” of task completions followed by relatively long periods of inactivity, were not present in the situation of the municipality.

A final limitation that needs to be mentioned is that only a restricted period (half a year) was used as a time window for the evaluation of the invoice handling’s process performance. We attempted to counter this issue with carrying out our analyses on the level of sub-logs as well, but we cannot rule out entirely that we have witnessed a temporary effect.

Implications for practical use. The most important implication from our work for practice is that workflow technology should not be assumed to level all geographical barriers between people just by itself. High expectations on workflow technology need to be re-adjusted, for example, when they are considered as infrastructure for worldwide operating enterprises that “follow the sun”. Explicit efforts must be taken to create equal circumstances for all involved workers if equal performance is desired. In addition, geographical proximity of workers favors their interaction, as was already suggested by the work of Allen [4]. Organizations must explicitly look for and implement procedures, tools, and housing opportunities to stimulate interaction patterns among actors or should expect differences in performance to occur.

5 Conclusions

WfMSs are supposed to efficiently and effectively support actors in the execution of business processes they are involved in, regardless of their geographical location. In this paper, we critically evaluated this assumption through a study into the performance of a WfMS in an organizational context. We analyzed a large workflow process logs with the ProM framework and associated tools. We found that the geographical location and distance between actors was a major distinguishing factor in process performance. The feedback from the organization brought a partial explanation for the phenomenon, i.e. that people are more inclined to urge others to complete their work when they are geographically close to them. Also, the positive effects of spontaneous interactions between collocated workers may be at work here.

Our paper contributes to a better understanding of the organizational effectiveness of workflow technology, which is an important but not so widely researched topic. Furthermore, the paper clearly demonstrates the feasibility of process mining techniques in evaluating current situations or answering managerial questions related to process enactment. Since only logs from a single organization were used, our results are clearly open to discussion.

In future work, we plan to repeat our analysis with logs from other organizations, taking into account other potential factors affecting performance (e.g. organizational hierarchy). It would be highly desirable to find organizations with highly distributed actors for reasons of comparison.

Acknowledgement

This research is supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Dutch Ministry of Economic Affairs. Byungduk Jeong is visiting the Department of Technology Management at Eindhoven University of Technology, supported by the Korea Research Foundation Grant founded by the Korean Government (KRF-2006-612-D00098). Moreover, we would like to thank the many people involved in the development of ProM.

References

1. van der Aalst, W.M.P., van Hee, K.M.: *Workflow Management: Models, Methods, and Systems*. MIT Press, Cambridge, MA (2002)
2. van der Aalst, W.M.P., Reijers, H.A., Song, M.: Discovering social networks from event logs. *Computer Supported Cooperative Work* 14(6), 549–593 (2005)
3. van der Aalst, W.M.P., Weijters, A.J.M.M., Maruster, L.: Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering* 16(9), 1128–1142 (2004)
4. Allen, T.: *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organisation*. MIT Press, Cambridge, MA (1977)
5. Barker, R.G.: *Ecological psychology*. Stanford University Press, Stanford, Calif (1968)
6. Becker, J., Vossen, G.: Geschäftsprozessmodellierung und Workflow-Management: Eine Einführung, chapter Geschäftsprozessmodellierung und Workflow-Management. Modelle, Methoden, Werkzeuge, Albany, pp. 17–26 (1996)
7. Boutellier, R., Gassmann, O., Macho, H., Roux, M.: Management of Dispersed R&D Teams. *R&D Management* 28(1), 13–25 (1998)
8. Carmel, E., Agarwal, R.: Tactical approaches for alleviating distance in global software development. *Software, IEEE* 18(2), 22–29 (2001)
9. Davenport, T.H.: *Process innovation*. Harvard Business School Press, Boston, Mass (1994)
10. van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M., van der Aalst, W.M.P.: The ProM framework: A new era in process mining tool support. In: Ciardo, G., Darondeau, P. (eds.) *ICATPN 2005*. LNCS, vol. 3536, pp. 444–454. Springer, Heidelberg (2005)
11. Dumas, M., van der Aalst, W.M.P., ter Hofstede, A.H.: *Process-aware information systems: bridging people and software through process technology* (2005)
12. Georgakopoulos, D., Hornick, M., Sheth, A.: An overview of workflow management: From process modeling to workflow automation infrastructure. *Distributed and Parallel Databases* 3(2), 119–153 (1995)

13. Grefen, P., Pernici, B., Sánchez, G.: Database Support for Workflow Management: The Wide Project. Kluwer Academic Pub., Dordrecht (1999)
14. Herbst, J., Karagiannis, D.: Workflow mining with InWoLvE. *Computers in Industry* 53(3), 245–264 (2004)
15. Herrmann, T., Hoffmann, M.: The Metamorphoses of Workflow Projects in their Early Stages. *Computer Supported Cooperative Work (CSCW)* 14(5), 399–432 (2005)
16. IDS Scheer. ARIS Process Performance Manager (ARIS PPM) (2002), <http://www.ids-scheer.com>
17. Kraut, R.E., Fussell, S.R., Brennan, S.E., Siegel, J.: Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. *Distributed work*, pp. 137–162 (2002)
18. Kueng, P.: The effects of workflow systems on organizations: A qualitative study. In: Van der Aalst, W.M.P., Desel, J., Oberweis, A. (eds.) *Business process management: Models, techniques, and empirical studies*, pp. 301–316. Springer, Berlin (2000)
19. Lawrence, P. (ed): *Workflow Handbook 1997*, Workflow Management Coalition. John Wiley and Sons, New York (1997)
20. Mohan, C., Alonso, G., Gunthor, R., Kamath, M.: Exotica: A Research Perspective on Workflow Management Systems. *Data Engineering Bulletin* 18(1), 19–26 (1995)
21. Oba, M., Onada, S., Komoda, N.: 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. IEEE, Washington, DC (2000)
22. Reijers, H.A., van der Aalst, W.M.P.: The Effectiveness of Workflow Management Systems: Predictions and Lessons Learned. *International Journal of Information Management* 56(5), 457–471 (2005)
23. Sengupta, K., Zhao, J.L.: Improving the communicational effectiveness of virtual organizations through workflow automation. *International Journal of Electronic Commerce* 3(1), 49–69 (1998)
24. Steinfield, C., Jang, C.Y., Pfaff, B.: Supporting virtual team collaboration: the TeamSCOPE system. In: *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, pp. 81–90 (1999)
25. Townsend, A.M., DeMarie, S.M., Hendrickson, A.R.: Virtual teams: Technology and the workplace of the future. *IEEE Engineering Management Review* 28(2), 69–80 (2000)