

Process Mining of Five Dutch Municipalities' Building Permit Application Process: The Value Added in E-Government

The BPI Challenge 2015

Hyeoung Seok Choi¹, Won Min Lee¹, Ye Ji Kim¹, Jung Hoon Lee¹, Chun Hoe Kim¹, Yu Lim Kang¹, Na Rae Jung¹, Seung Yun Kim¹, Eui Jin Jung¹, Na Hyeon Kim¹

APPS(Advanced Process Performance reSearch center), Department of MIS, School of Business Administration, Myongji Univ., Seoul, Korea
{choihs91, nu9177, a92344349, cft0714, ggochun1030, lim10046, wjdskf094, tmddus914, tyghvb1210, skgus978}@gmail.com

Abstract. An increasing number of countries have been making sizable investments in e-government to achieve efficiency in national administration and improvement of satisfaction in nationwide services. Process mining can create new insights for e-government initiatives by analyzing event logs stored in e-government IT systems. To validate the added value of process mining in e-government, we analyze real life event logs of five Dutch municipalities' building permit application process by using process mining and other analytical techniques. To fully understand and analyze the logs, we used a variety of data analysis techniques and powerful tools such as Disco, Weka, Oracle DBMS. We offer evidence-based answers to the questions and demonstrate the added value of process mining in e-government. Finally, concluding remarks and recommendations for improvements are discussed.

1. Introduction

There are some countries, with high e-Government (Electronic Government) Development Index of UN such as South Korea, Netherlands, Australia, United States of America, United Kingdom, which have striven for achieving efficiency in national administration and improvement of satisfaction in nationwide services, by making sizable investments in IT [1]. The government organizations of such countries have stored the vast amount of the event logs to be utilized in process mining analysis, for they are currently performing their administrative works based on their self-developed IT systems or packaged applications [2]. Utilizing and analyzing such event logs can be an enormous help in e-Government development efforts around the world, which will be able to contribute to the efficiency of the administration of state and public service satisfaction.

¹ Corresponding author. Tel.: +82 2 300 0776.

Then how can we analyze and utilize the enormous event logs each of the government organizations has accumulated? Process Mining can support cost cutting, waste elimination, best practice application, value innovation, etc. by utilizing the vast amount of process data recorded in an IT system which supports administrative task performances of the government organizations [3]. A scientific analysis and improvement method utilizing process mining can help to overcome some limitations of existing traditional methods proceeded manually, such as interview, workshop, etc. Based on process mining, we intend to draw some analysis results by visualizing relevant process. Such dataset includes data related to every building permit application process performed by the five municipalities in Dutch for the past 4 years. We tried to draw the analysis results for six questions (Question 1~6) which the BPI Challenge raised. The question 0, we raised, is about basic assumptions for drawing analysis results for Question 1~6.

- **Question 0:** How we could draw the Reference Process Model of the Building Permit Application Process of five municipalities in Dutch and what is the meaning of 'process stage'?
- **Question 1:** What are the possible points for improvement on the organizational structure for each of the municipalities?
- **Question 2:** Where are differences in throughput times between the municipalities and how can these be explained?
- **Question 3:** The employees of two of the five municipalities have physically moved into the same location recently. Did this lead to a change in the processes and if so, what is different?
- **Question 4:** What are the roles of the people involved in the various stages of the process and how do these roles differ across municipalities?
- **Question 5:** What are the differences in control flow between the municipalities?
- **Question 6:** Some of the procedures will be outsourced from 2018, i.e. they will be removed from the process and the applicant needs to have these activities performed by an external party before submitting the application. What will be the effect of this on the organizational structures in the five municipalities?

We believe that the scientific answers to the questions will be helpful in the improvement of the building permit application process of the municipalities of Dutch. In addition, our analysis results will be able to offer many suggestions for the government organizations around the world, which intend to promote efficiency of the administrative tasks and raise the nationwide services through vast investment in the e-Government initiatives.

2. Understanding of the Data

As mentioned before, five Dutch municipality datasets of the building permit applications were provided to us. Prior to the in-depth analysis, we needed to understand the data basically. To understand the datasets, therefore, we imported them into Oracle DBMS. As the result of that process, the summarized information

about the attributes of each dataset is provided in <Fig. 1-5>. Based on this understanding of the data, we have created relevant datasets to address the questions.

Municipality 1		
Case	1,199	case
Event	52,217	case
Activity	289	Type
Resource	23	Type
CompleteTimestamp	2010.10.05 07:00:00/ 2015.08.01 07:00:00	Min/Max
(case) IDofConceptCase	811 (Including "null")	Type
(case) Includes_subCases	3 (Including "null")	Type
(case) Responsible_actor	21 (Including "null")	Type
(case) SUMleges	-2234.54/104549.7 (Including "null")	Min/Max
(case) caseProcedure	3 (Including "null")	Type
(case) caseStatus	3	Type
(case) case_type	1	Type
(case) landRegisterID	141 (Including "null")	Type
(case) last_phase	27	Type
(case) parts	89 (Including "null")	Type
(case) requestComplete	2	Type
(case) termName	14 (Including "null")	Type
action_code	398 (Including "null")	Type
concept.name	398	Type
lifecycle:transition	1	Type
monitoringResource	26	Type
planned	2010-10- 02T13:04:41+02:00/ 2015-03-04T13:12:17+01:00	Min/Max
question	422	Type

Fig. 1. Attributes of 'Municipality 1 (log1)' Dataset.

Municipality 2		
Case	832	case
Event	44,354	case
Activity	304	Type
Resource	11	Type
CompleteTimestamp	2010.06.29 07:00:00/ 2015.03.04 08:00:00	Min/Max
(case) IDofConceptCase	311 (Including "null")	Type
(case) Includes_subCases	3 (Including "null")	Type
(case) Responsible_actor	7	Type
(case) SUMleges	-8384.015/52939.727 (Including "null")	Min/Max
(case) caseProcedure	3 (Including "null")	Type
(case) caseStatus	3 (Including "null")	Type
(case) case_type	1	Type
(case) landRegisterID	204 (Including "null")	Type
(case) last_phase	23 (Including "null")	Type
(case) parts	64 (Including "null")	Type
(case) requestComplete	2	Type
(case) termName	10 (Including "null")	Type
action_code	409 (Including "null")	Type
concept.name	410	Type
lifecycle:transition	1	Type
monitoringResource	9	Type
planned	2010-10- 13T11:57:32+02:00/ 2015-03-05T11:06:51+01:00	Min/Max
question	391	Type

Fig. 2. Attributes of 'Municipality 2 (log2)' Dataset.

Municipality 3		
Case	1,409	case
Event	59,681	case
Activity	277	Type
Resource	14	Type
CompleteTimestamp	2010.01.01 08:00:00/ 2015.03.05 08:00:00	Min/Max
(case) IDofConceptCase	805 (Including "null")	Type
(case) Includes_subCases	3 (Including "null")	Type
(case) Responsible_actor	20	Type
(case) SUMleges	-6.744/129309.03 (Including "null")	Min/Max
(case) caseProcedure	3 (Including "null")	Type
(case) caseStatus	2	Type
(case) case_type	1	Type
(case) landRegisterID	207 (Including "null")	Type
(case) last_phase	27	Type
(case) parts	94 (Including "null")	Type
(case) requestComplete	2	Type
(case) termName	14 (Including "null")	Type
action_code	383	Type
concept.name	383	Type
lifecycle:transition	1	Type
monitoringResource	22	Type
planned	2010-10- 09T12:00:06+02:00/ 2015-03-05T15:28:08+01:00	Min/Max
question	540	Type

Fig. 3. Attributes of 'Municipality 3 (log3)' Dataset.

Municipality 4		
Case	1,053	case
Event	47,293	case
Activity	272	Type
Resource	10	Type
CompleteTimestamp	2009.11.18 08:00:00/ 2015.03.05 08:00:00	Min/Max
(case) IDofConceptCase	622 (Including "null")	Type
(case) Includes_subCases	3 (Including "null")	Type
(case) Responsible_actor	9 (Including "null")	Type
(case) SUMleges	-1970.5125/42058.324 (Including "null")	Min/Max
(case) caseProcedure	3 (Including "null")	Type
(case) caseStatus	2	Type
(case) case_type	1	Type
(case) landRegisterID	86 (Including "null")	Type
(case) last_phase	25	Type
(case) parts	89 (Including "null")	Type
(case) requestComplete	2	Type
(case) termName	14 (Including "null")	Type
action_code	355 (Including "null")	Type
concept.name	356	Type
lifecycle:transition	1	Type
monitoringResource	12	Type
planned	2010-10- 08T10:18:07+02:00/ 2015-03-05T13:57:02+01:00	Min/Max
question	374	Type

Fig. 4. Attributes of 'Municipality 4 (log4)' Dataset.

Municipality 5		
Case	1,156	case
Event	59,083	case
Activity	285	Type
Resource	22	Type
CompleteTimestamp	2009.11.23 08:00:00/ 2015.03.03 23:31:43	Min/Max
(case) IDofConceptCase	638 (Including "null")	Type
(case) Includes_subCases	3 (Including "null")	Type
(case) Responsible_actor	8 (Including "null")	Type
(case) SUMleges	0/227407.66 (Including "null")	Min/Max
(case) caseProcedure	3 (Including "null")	Type
(case) caseStatus	3	Type
(case) case_type	1	Type
(case) landRegisterID	857 (Including "null")	Type
(case) last_phase	25	Type
(case) parts	107 (Including "null")	Type
(case) requestComplete	2	Type
(case) termName	11 (Including "null")	Type
action_code	386 (Including "null")	Type
concept.name	389	Type
lifecycle.transition	1	Type
monitoringResource	16	Type
planned	2010-10-13T11:32:06+02:00/ 2015-03-10T15:31:43+01:00 (Including "null")	Min/Max
question	625	Type

Fig. 5. Attributes of 'Municipality 5 (log5)' Dataset.

3. Analysis from of the Questions

3.1 Question 0: The Primary Definition for Understanding the Building Permit Application Process

3.1.1 Definition of Completed Cases

Since the data set given to us for analysis was located in certain period, it is natural for the data set to contain some incomplete cases. We understood that these incomplete cases can cause some analytical errors related to statistics, performance duration, etc. We, therefore, needed to define completed case prior to our analysis, and we decided to regard cases containing 'close case' activities – meaning that the case was finished/completed - as 'completed case'. When the definition applied, the frequency of complete cases per each five municipality is listed in <Fig. 6 >.

	A: total case (#)	B: completed case (#)	=B/A(%)
Municipality 1	1,199	171	14.26%
Municipality 2	832	614	73.80%
Municipality 3	1,409	53	3.76%
Municipality 4	1,053	98	9.31%
Municipality 5	1,156	797	68.94%
Total	5,649	1,733	30.68%

Fig. 6. completed cases of each municipality.

3.1.2 Definition of Process Stage

The given data set from the building permit application process consists of main process and several sub processes. We distinguished those processes by using the action_code columns formed as ‘##_characters_####’ and concept:name columns. And we also referred the first two digits and the following characters of the action_code (as ‘##_characters’), excluding the last three digit telling the order of the process, for distinguishing those processes. As the result, we figured out that every five municipality consists of one (1) main process and 28 sub processes. We regarded each relevant process as one step of the process, so we named each of them as ‘process stage’. (Cf. <Fig. 7>)

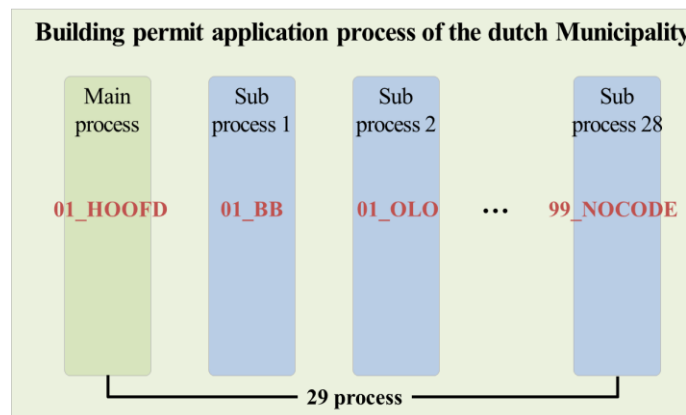


Fig. 7. Process Composition of the Building Permit Application Process in Dutch

3.1.3 Definition of Reference Process Model

The Reference Process Model reflects some rules or policies which are regulated legally and/or in policy and describes some effective and efficient way of operating duties. That is, it is sort of a manual for process model, which describes the proper order of the process. However, the process sometimes takes a different sequence even though the work flow of actual performance applies the reference process model. We, therefore, have tried to understand the building permit application process, prior to grasp the workflow, to get a reference process model. To get the reference process

model, we used the column, ‘action_code’, which contains some information related to the progress stages and the order of performance of them.

Before getting the reference process model, we clarified if there is one reference process model covering all of the five municipalities. So we examined whether the corresponding relation between action_code (i.e. process) and activity applies to the five municipalities as same. And we finally found that there are several activities for each action_code. Moreover, the correspondence between activity and action_code are same in the five municipalities when we exclude some unperformed activities (Cf. <Fig. 8>). We, therefore, based on the fact that each process and its activities are found in every municipality as same, concluded there is one reference process model.

ACTION_CODE	Activity				
	Municipality 1	Municipality 2	Municipality 3	Municipality 4	Municipality 5
01_HOOFD_010	register submission date request	register submission date request	register submission date request	register submission date request	register submission date request
01_HOOFD_011	OLO messaging active	OLO messaging active	OLO messaging active	OLO messaging active	OLO messaging active
01_HOOFD_012		application submitted through OLO	application submitted through OLO	application submitted through OLO	application submitted through OLO
01_HOOFD_015	phase application received	phase application received	phase application received	phase application received	phase application received
01_HOOFD_020	reception through OLO	reception through OLO	reception through OLO	reception through OLO	reception through OLO
01_HOOFD_020	send confirmation receipt	send confirmation receipt	send confirmation receipt	send confirmation receipt	send confirmation receipt
01_HOOFD_030_1	send confirmation receipt	send confirmation receipt	send confirmation receipt	send confirmation receipt	send confirmation receipt
01_HOOFD_030_2	enter senddate acknowledgement	enter senddate acknowledgement	enter senddate acknowledgement	enter senddate acknowledgement	enter senddate acknowledgement
01_HOOFD_040	forward to the competent authority	forward to the competent authority	forward to the competent authority	forward to the competent authority	forward to the competent authority
01_HOOFD_050	inform BAG administrator	inform BAG administrator	inform BAG administrator	inform BAG administrator	inform BAG administrator
01_HOOFD_050	applicant is stakeholder	applicant is stakeholder	applicant is stakeholder	applicant is stakeholder	applicant is stakeholder
01_HOOFD_055	temporary permit		temporary permit	temporary permit	
01_HOOFD_055	investigate BAG objects	investigate BAG objects	investigate BAG objects		investigate BAG objects
01_HOOFD_060	regular procedure without MER	regular procedure without MER	regular procedure without MER	regular procedure without MER	regular procedure without MER
01_HOOFD_061	start WABOprocedure	start WABOprocedure	start WABOprocedure	start WABOprocedure	start WABOprocedure
01_HOOFD_065_0	create procedure confirmation	create procedure confirmation	create procedure confirmation	create procedure confirmation	create procedure confirmation
01_HOOFD_065_1	send procedure confirmation	send procedure confirmation	send procedure confirmation	send procedure confirmation	send procedure confirmation
01_HOOFD_065_2	enter senddate procedure confirmation	enter senddate procedure confirmation	enter senddate procedure confirmation	enter senddate procedure confirmation	enter senddate procedure confirmation
01_HOOFD_080	date for inspection MER	date for inspection MER	date for inspection MER	date for inspection MER	date for inspection MER

Fig. 8. Some Activities Corresponding Action_code in the Five Municipalities

We got the reference process model with understanding the first two digits and the following characters of an action_code as a process stage and the last three digits of it as the order of the performance. There are 28 processes consist of its 500 action_codes for the reference process model. <Data of Reference Process Model is attached.>

3.2 Question 1: What are the possible points for improvement on the organizational structure for each of the municipalities?

3.2.1 Understanding the Question

To reduce accidental or intentional system abuse, the management needs to segregation duties not compatible together and operate them individually. Also, if the segregation of duties is not possible because of the lack of human resources in the organization, they should apply the practical rule as possible. Alternatively, another means of controlling could be requested; such as activity monitoring, activity assessment and managing/supervising.

Those 5 municipalities separately secured one relevant responsible actor ((case) Responsible_actor Column) and activity operators (Resource Column) and activity supervisors (monitoringResource Column) for each activity. Although those three roles need to be separated as possible, one person sometimes carries out more than one role for the lack of human resources. If such circumstance occurs inevitably

because of the lack of human resource, at least the activity supervisor need stick to his/her role. As a solution to this question, we are, accordingly, intended to suggest some measures to improve the organizational structure for each of the municipalities for reducing corruption(obliquity). The question 0, we raised, is about basic assumptions for drawing analysis results for Question 1~6

3.2.2 Key Data Columns for Addressing the Question 1

- **Resource:** Activity Operator
- **(case) Responsible_actor:** Person in charge of the relevant case(process instance)
- **monitoringResource:** A supervisor monitoring resource (HR) to perform activity without any problem.

3.2.3 Levels of Analysis and the Data for Analysis

A responsible actor ((case) Responsible_actor Column) has the same value for one case. That is, the (case) Responsible_actor Column is a case based column. However, it is recorded differently for each activity in regard of activity operators (Resource Field) and activity supervisors (monitoringResource Field). So, in this analysis, we set all events recorded in the log as subjects to analysis, regardless if they are completed or not.

Some events, however, of which at least one column out of those three has null value is null and which doesn't meet the condition for analysis have been excluded from the analysis. As a result, 19 events without (case) Responsible_actor Value in the log 1 have been excluded and following events are finally selected as subjects to analysis: [log1: 52,198 events], [log2: 44,354 events], [log3: 59,681 events], [log4: 47,293 events], [log5: 59,083 events].

3.2.4 Analysis

We analyzed the level of compliance of the Segregation of Duties (SoD) of 5 municipalities by utilizing the selected data. Based on the result of the analysis, according to the level of segregation of duties, we classified 5 municipalities into the following categories: best, good, bad, worst (Cf. <Fig. 9>)

No.	Resource	(case) Responsible actor	monitoring Resource	SoD level
1	A	B	C	Best
2	A	A	B	Good
3	A	B	B	Bad
4	A	B	A	
5	A	A	A	Worst

Fig. 9. Standard of classification of Segregation of Duties level (SoD level)

1. **‘Best’** level of the Segregation of Duties level (SoD level) indicates that all of the duties –Resource (regarded as activity operators), (case) Responsible_actor (regarded as a responsible actor) and monitoringResource (regarded as activity supervisors)- are performed by different resources.
2. **‘Good’** level of the Segregation of Duties level (SoD level) indicates that some of the duties –Resource (regarded as activity operators) and (case) Responsible_actor (regarded as a responsible actor)- are performed by same resource and monitoringResource (regarded as activity supervisors) by a different resource.
3. **‘Bad’** level of the Segregation of Duties level (SoD level) indicates that some of the duties - (case) Responsible_actor (regarded as a responsible actor) and monitoringResource (regarded as activity supervisors) - are performed by same resource and Resource (regarded as activity operators) by a different resource.
4. **‘Bad’** level of the Segregation of Duties level (SoD level) indicates that some of the duties –Resource (regarded as activity operators) and monitoringResource (regarded as activity supervisors)- are performed by same resource and (case) Responsible_actor (regarded as a responsible actor) by a different resource.
5. **‘Worst’** level of the Segregation of Duties level (SoD level) indicates that all of the duties –Resource (regarded as activity operators), (case) Responsible_actor (regarded as a responsible actor), monitoringResource (regarded as activity supervisors) - are performed by a same resource.

3.2.5 Analysis Results

<Fig. 10> is a summary of the results of the analysis, which is based on the standard of <Fig. 9> regarding to the Segregation of Duties level (SoD level) of 5 municipalities. Even though there are some differences among those 5 municipalities, generally the ‘Bad’ and ‘Worst’ level take higher percentage. In case of the municipality 1, ‘Best’ level showing three separated duties is the highest at 36.64%, which is much higher than the average value (10.38%) of all municipalities and the ‘Worst’ level was very low by 5.95%. The percentage of ‘Worst’ level of the municipality 5 is the highest as 51.21%. ‘Bad’ level is top ranked for all of the 5 municipalities.

	Best		Good		Bad		Worst		Total
	#	%	#	%	#	%	#	%	#
Municipalty 1	19,124	36.64	981	1.88	28,989	55.54	3,104	5.95	52,198
Municipalty 2	1,307	2.95	789	1.78	24,535	55.32	17,723	39.96	44,354
Municipalty 3	4,743	7.95	3,281	5.50	30,042	50.34	21,615	36.22	59,681
Municipalty 4	637	1.35	196	0.41	30,884	65.30	15,576	32.94	47,293
Municipalty 5	1,436	2.43	467	0.79	26,922	45.57	30,258	51.21	59,083
Total	27,247	10.38	5,714	2.18	141,372	53.83	88,276	33.61	262,609

Fig. 10. Summary of the Analysis Results regarding to the Segregation of Duties level (SoD level) of 5 municipalities

There are some possible points for improvement that we suggest based on the analysis results. First, readjustment of Role and Responsibility (R&R) is required to modify the percentage of the 'Bad' level of those 5 municipalities. Activity supervisors don't need to perform activities of activity operators or a responsible actor multiply in one case. Second, we have figured out that the level of segregation of duties is clearly different for each municipality even though there are similar numbers of resources in 5 municipalities. The rest of other municipalities, therefore, need to understand and follow the organizational structure and the way to perform the duties of the 'municipality 1'.

3.3 Question 2: Where are differences in throughput times between the municipalities and how can these be explained?

3.3.1 Understanding the Question

It has been found that there are some differences in throughput times of processes, the subject of our analysis, between municipalities. For this question, therefore, we intend to figure out some meaningful parameters to explain the differences. Also, based on such parameters, we are going to show how the throughput times of each process are different between municipalities.

3.3.2 Selecting the Suitable Cases for the Question

Before proceeding the analysis, we need to rule out some incompleting cases which may cause some errors in the actual throughput time measurement. we selected only completed cases of 5 municipalities as the subject to our analysis (cf. paragraph 3.1.1).

3.3.3 Analysis

3.3.3.1 A Summary of Methods of Analysis

To find meaningful parameters, we generated a separate ASSESSMENT column for evaluating the throughput times of processes. We arranged those, 1,733 completed cases of 5 municipalities in ascending order in accordance with the throughput times to set standards for evaluation. Then we classified those top 25% (433) cases as 'Short', the bottom 25% (433) cases as 'Long' and medium 50% (867) cases as 'Average'. For understanding definite differences, we excluded those Average (867) cases and made those Long and Short (866) cases the total subject for our analysis. Applying Decision Tree technique to the selected subjects of analysis, we found meaningful parameters.

3.3.3.2 Candidate Parameters for Decision Tree Analysis

A specific selection stage for parameters which are possible to affect process throughput times was necessary for us to find meaningful parameters. Therefore, following 4 parameters was selected as alternative parameters for Decision Tree analysis among all of those columns of data set.

- **#_events:** Event number of relevant cases from the beginning through the completion.
- **(case) Includes_subCases:** Whether relevant case includes any serve cases(J: Yes, N: No)
- **(case) Responsible_actor:** Person in Charge of process instances.
- **(case) caseProcedure:** Whether procedures of relevant case are performed 'extensively' or 'regularly'(Uitgebreid: extensive, Regulier: regular)

3.3.3.3 Process Throughput Times of Each Municipality

We intend to analyze several causes of the different throughput times by applying the candidate parameters we found on 4.6.3.2 to subject cases. To minimize possible errors caused by widely different numbers of cases among municipalities, we are planning to group all municipalities by similar completed case numbers and compare them.

- Group A: Municipality 2, 5
- Group B: Municipality 1, 3, 4

3.3.4 Analysis Results

3.3.4.1 Parameters

<Fig. 11.> is a summary of the results of analysis of alternative parameters. It was decided that only '(case) caseProcedure' among those 4 parameters affects the process throughput times. The credibility of the parameter is slightly low as 67.21%. However, there should be a consideration that many cases are distributed throughout some Null value, without any exact value. Also, when the case procedure performance ((case) caseProcedure Column) is extensive(Uitgebreid), there is 92.35% of possibility of this procedure to be in the bottom 25%.

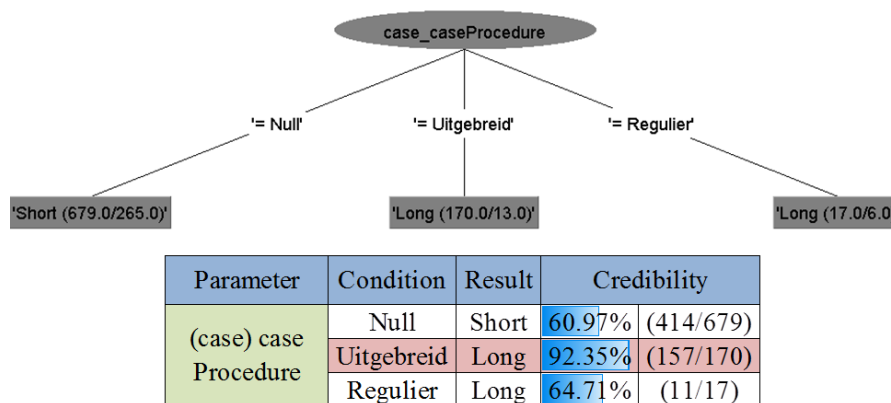


Fig. 11. Alternative Parameters Analysis Results (Data Mining tool - weka)

3.3.4.2 Differences between Process Throughput times of each municipality

<Fig. 12> summarizes the result of the analysis of process throughput time of each municipality by using (case) caseProcedure, a meaningful parameter.

- **Group A [Municipality 2, 5]**
In case of Municipality 2 which has longer throughput time, it has a bigger ratio of extensive (Uitgebreid) case performance procedure.
- **Group B [Municipality 1, 3, 4]**
In case of Municipality 4 which has long throughput time, it has a bigger ratio of extensive (Uitgebreid) case performance procedure than of Municipality 1 and 3.

		Municipality 1		Municipality 2		Municipality 3		Municipality 4		Municipality 5	
		#	%	#	%	#	%	#	%	#	%
#_cases		171		614		53		98		797	
Duration (day)	Average	96.6		176.4		93.8		123.9		109.2	
	Median	167.3		123.2		69.4		95.9		87.5	
(case) case Procedure	Uitgebreid	22	12.87%	113	18.40%	7	13.21%	20	20.41%	50	6.27%
	Regulier	0	0.00%	40	6.51%	8	15.09%	0	0.00%	7	0.88%
	null	149	87.13%	461	75.08%	38	71.70%	78	79.59%	740	92.85%

Fig. 12. Analysis between Throughput Times Of Municipalities.

In conclusion of the analysis discussed above, we figured out that all cases of extensive performance procedure generally have longer throughput time.

3.4 Question 3: The employees of two of the five municipalities have physically moved into the same location recently. Did this lead to a change in the processes and if so, what is different?

3.4.1 Understanding the Question

The assignment of human resources directly affect process efficiency and achievements of whole organization [4]. We would like to grasp what kind of changes in places of duty caused in process efficiency of municipalities and achievements of whole organization. We also intend to analyze the main cause of the changes. Our analysis will be based on the following hypotheses:

1. The newly transferred employee affects other cases as well as the specific cases which he/she is directly involved in.
2. We don't consider any interventions of other circumstantial factors (such as economic growth, government policy, etc.), except the transfer of relevant Resources.

3.4.2 Levels of Analysis and the Data for Analysis

Including incomplete cases can cause some errors related to statistics and/or duration, etc. in process mining analysis. We, therefore, only selected fully-completed cases as subjects to this analysis. Also, comparing all of the processes, there were some difficulties in analyzing the differences from each process because of the huge

quantity of activities. We have concentrated this analysis on the main process (action_code: '01_HOOFD_###') and the related events to overcome such limitations.

3.4.3 Analysis

To compare the processes before and after the change in a place of duty of Resource (i.e. an employee), it was necessary for us to find the particular Resource, a subject to the movement, and the time of transfer. When specific Resources transferred to other municipalities, the relevant data of such Resources are found in the datasets of multiple municipalities. '560530' and '560532' Resources are originally belong to 'municipality 2', but they are also found in the dataset of 'municipality 5'. '560752' and '560849' Resources also seem to transfer from 'municipality 4' to 'municipality 5'.

We analyzed the time when the Resources appear in datasets to figure out when the Resources transferred (Cf. <Fig.13>). As the result of the analysis, the time when '560752' appears first in Municipality 5 is Mar, 15, 2012. However, '560752' had performed only one event for about 200 days until Oct, 01, 2012. Based on the information, we consider the date '2012.10.01' on which '560752' performed an event at the time of its transfer to 'municipality 5'. And using similar method, we found the time of transfer of another three Resources. '560530' showed the latest time of transfer, May, 14, 2014, among the four employees.

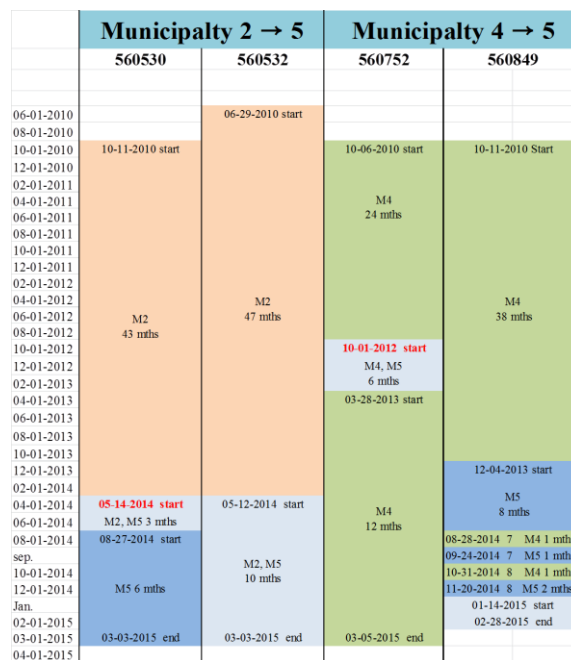


Fig. 13. Transference Time of Four Reassigned Employees

To find changes in process related to those four reassigned employees, appropriate process models should be found. As seen in <Fig. 14>, we organized periods of each dataset to be used for finding process models. First, the period prior to Oct, 01, 2012 was designated as **‘Period without moving’**, which means there was no change in place of duty of any employee. The next period, **‘Period influenced by moving’**, shows it affected by the changes in places of duty. Also, to figure out clearly the changes of process affected by the transfer of new employees to ‘Municipality 5’, we subdivided those two periods mentioned above. The **‘Period influenced by moving’** is separated into the period before the transfer of employees was finished (**Transition period**: Oct, 01, 2012 – May, 31, 2014) and the period that all employees are assigned (**After period**: May, 14, 2014 ~). In the **‘Period without moving’**, 293 days (Dec, 12, 2011 – Sep, 30, 2012) which overlaps with the **‘After Period’** were set as **‘Before Period’**. For drawing more accurate conclusion, we tried to understand the changes in the process focusing on **‘Before period’** and **‘After period’**, excluding the **‘Transition period’**.

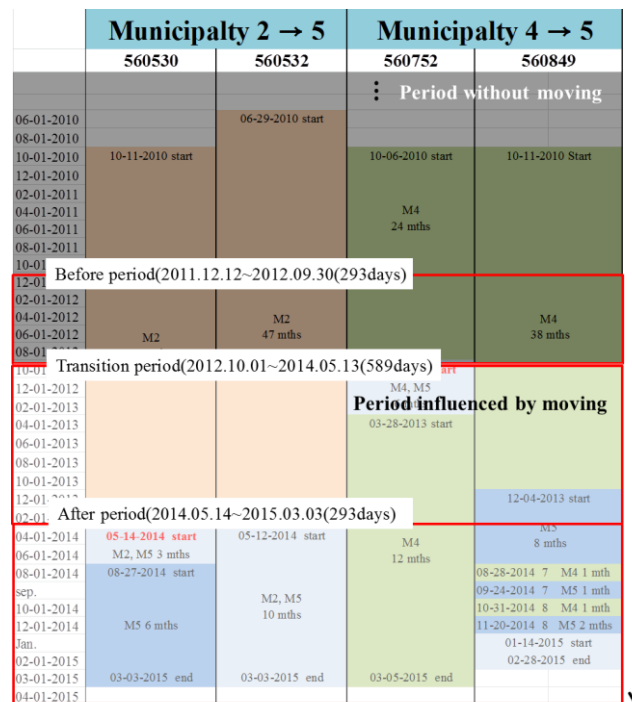


Fig. 14. Periods organized for Analyzing Changes in Process

3.4.4 Analysis results

Following the conclusions of the analysis, we found there are many differences between ‘Before period’ and ‘After period’ in operating activities of Municipality 5. Examining overall statistics of every process, 119 cases, 4,042 events, and 57 activities were performed during ‘Before period’. On the contrary, during ‘After

period’, there were 56 cases, 2,201 events, and 66 activities. There is a difference of 4 operating days between Before and After period (Before period: 85.4 days vs. After period: 89.2 days).

3.4.4.1 Differences in the Beginning Parts of Processes

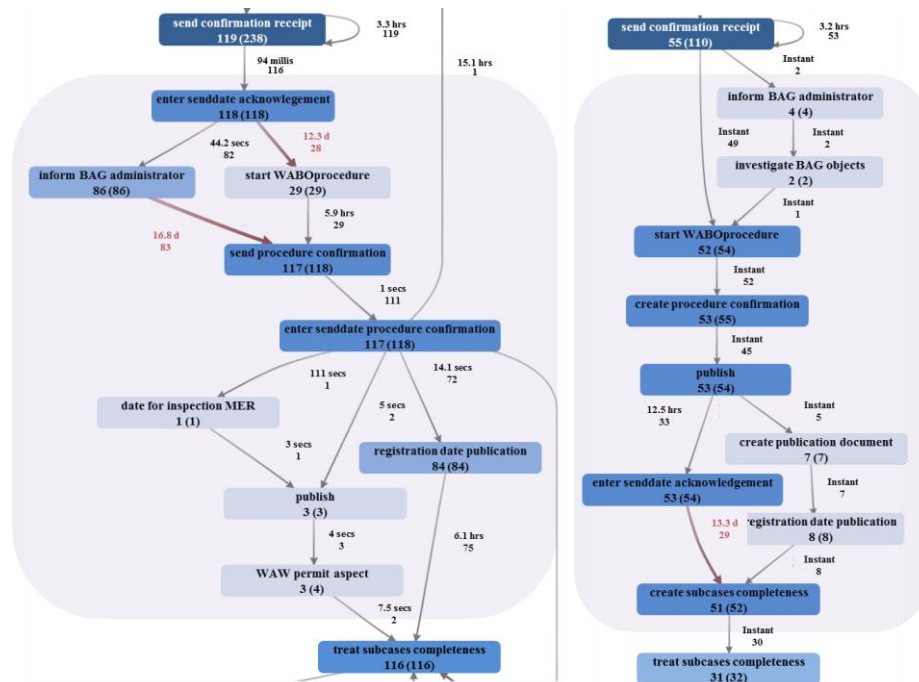


Fig. 15. Changes in Processes and the Difference of Mean Duration between ‘Before Period’ (left) and ‘After Period’ (right) 1 (Process Mining tool - Disco: Activities (100%), Paths (0%))

In the <Fig. 15 > which shows the beginning of processes, we found differences between processes of each ‘send confirmation receipt’ and ‘treat subcases completeness’ activity. There are 9 processes which are performed between these two activities. Then five new activities, including ‘investigate BAG objects’, are shown on the ‘After period’ process map. And there could be a lot of changes in the order of the existing performances of activities. For example, in case of ‘enter senddate acknowledgement’, it was performed right after ‘send confirmation receipt’ during the ‘Before period’ but, during the ‘After period’, it was performed in the latter part of ‘publish’ activity. There are also some differences in duration as well as the order of the process performances. First, we found the great decrease of the average operating time of ‘inform BAG administrator’. Going through the average duration from ‘send

confirmation receipt’ to ‘treat subcases completeness’, we figured out that it takes 14.75 days during the ‘After period’ although it takes 18.61 days during the ‘Before period’.

Based on the analysis results, we noticed there were many changes in processes between the activities ‘send confirmation receipt’ and ‘treat subcases completeness’, and the changes in processes shortened the duration by 3.86 days (20.74%) on average.

3.4.4.2 Differences in the Latter Parts of Processes

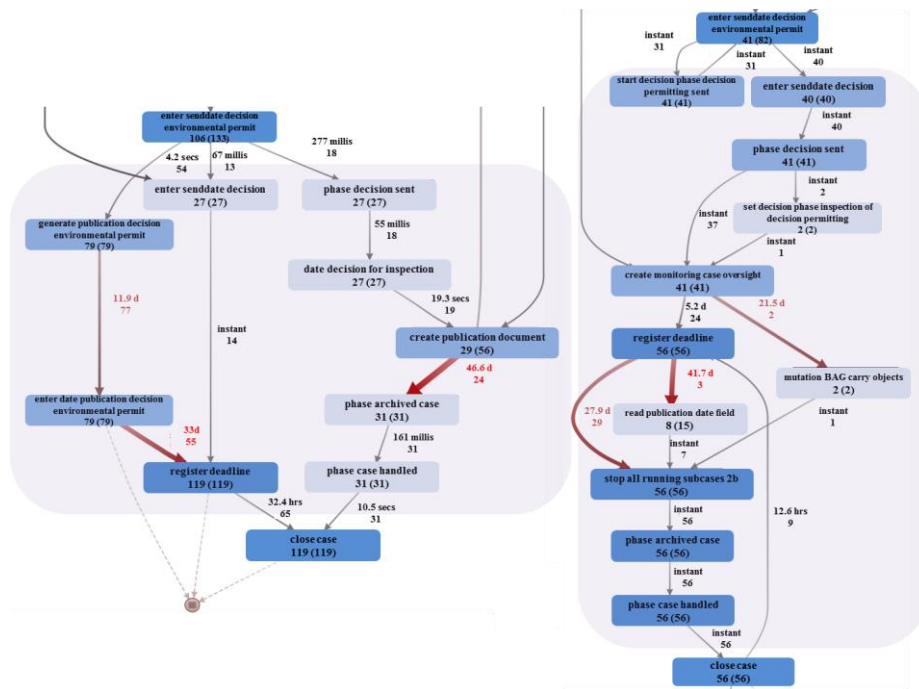


Fig. 16. Changes in Processes and the Difference of Mean Duration between ‘Before Period’ (left) and ‘After Period’ (right) 2 (Process Mining tool - Disco: Activities (100%), Paths (0%))

We have found some noticeable differences between 'enter senddate decision environmental permit' and 'close case' activities from <Fig. 16> referring to the latter parts of processes. During ‘Before period’, only 9 activities were performed, but the number of activities were increased by 11 during ‘After period’. Four previous activities vanished and 6 new activities such as ‘create monitoring case oversight’ were created. We also found some differences relating to duration. The average duration between ‘enter senddate decision environmental permit’ and ‘close case’ increased by 9.42 days (20.15%): Before period(46.74 days) vs. After period, (56.16 days).

3.5 Question 4: What are the roles of the people involved in the various stages of the process and how do these roles differ across municipalities?

3.5.1 Understanding the Question

There are thousands of business processes of any corporation which is found with business rules applied to. And it – business process – has been getting more complicated, forming hierarchy structures [5]. Each hierarchy structure generally has its separate business intent. And activities composing business related subprocesses are performed by activity operators of various roles. We assumed the main/subprocesses composing the building permit application process as various stages. With such assumption, we plan to find what are the roles of activity operators, involved in various stages of the process. We also look though the characteristics of each municipality by analyzing the differences of roles of the activity operators for each municipality.

3.5.2 Key Data Columns for Addressing Question 4

- **Activity:** Each activity belonging to the building permit application process.
- **Resource:** Process activity operator.
- **action_code:** As a form of ‘##_ABCD_###’, it refers to the level (character) and the order (the last three digits) of the processes. More than one activity corresponds to one action_code.
- **concept:name:** Has the same value with action_code, but the difference in a null value. The null value of action_code is recorded as ‘99_NOCODE_##’.

3.5.3 Levels of Analysis and the Data for Analysis

It was necessary to grasp to which process stage (i.e. main or sub process) the activities operated by the resources and relevant activities correspond for figuring out the roles of the activity operators. Therefore, in this analysis, we decided to make every event of all five municipalities the subject to this analysis to study all activities performed by the activity operators.

3.5.4 Analysis

There are more than one activity belonged to one action_code, and they – the action_code and the activities - show N:N relation. As we mentioned before, the main and sub processes of the target process were considered as process stage. Process stage and Action code refer to 1:N relation. (Cf. < Fig. 17>)

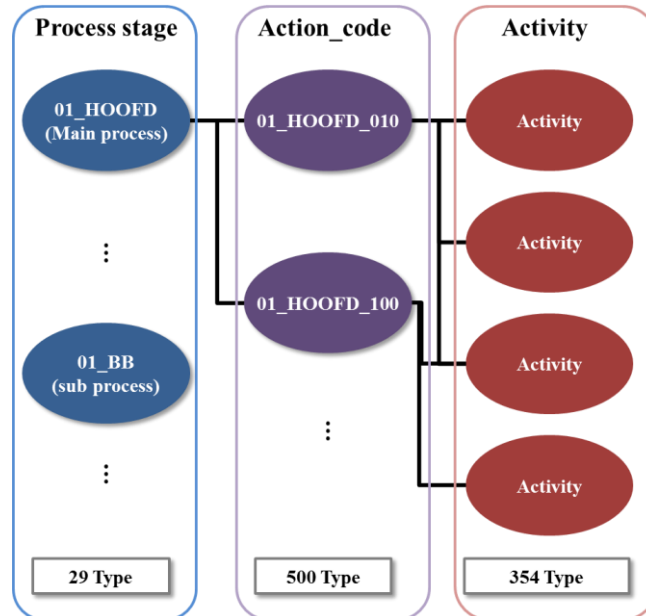


Fig. 17. Relation to Process Stage, Action_code, and Activity of Five Municipalities

We found there are differences in the number of process stages in which the activity operator (Resource) is involved for each municipality. (Cf. <Fig. 18>) We classified operators of each municipality who are involved in more than 80% of overall activities as ‘person involved in various stages’ and discovered which activities they operate. (Among the activity operators (Resources), ‘560594 (79.17%)’ and ‘560598 (79.17%)’, which are nearly 80%, are included in this analysis.)

Municipality 1 (20 process stages)			Municipality 2 (26 process stages)			Municipality 3 (23 process stages)			Municipality 4 (26 process stages)			Municipality 5 (24 process stages)		
Resource	# of process stages involved	%	Resource	# of process stages involved	%	Resource	# of process stages involved	%	Resource	# of process stages involved	%	Resource	# of process stages involved	%
2670601	19	95.00%	560458	24	92.31%	3442724	21	91.30%	1550894	23	88.46%	560604	23	95.83%
3273854	18	90.00%	560519	23	88.46%	560454	20	86.96%	560752	23	88.46%	560602	22	91.67%
560872	18	90.00%	4634935	21	80.77%	560749	20	86.96%	560821	23	88.46%	560429	21	87.50%
560462	17	85.00%	560521	21	80.77%	2013365	18	78.26%	560852	23	88.46%	560600	21	87.50%
560890	17	85.00%	560532	21	80.77%	560741	17	73.91%	560781	22	84.62%	560594	19	79.17%
560912	16	80.00%	560530	20	76.92%	560673	16	69.57%	560849	17	65.38%	560598	19	79.17%
560950	15	75.00%	20987361	8	30.77%	560696	16	69.57%	560812	11	42.31%	560608	18	75.00%
560894	14	70.00%	22445896	8	30.77%	5025869	10	43.48%	560431	6	23.08%	1254625	17	70.83%
11744364	11	55.00%	560598	8	30.77%	3122446	9	39.13%	560796	4	15.38%	560596	17	70.83%
560925	9	45.00%	560429	3	11.54%	560922	6	26.09%	6	2	7.69%	560530	10	41.67%
560589	8	40.00%	560528	3	11.54%	3069866	4	17.39%				560613	10	41.67%
560881	7	35.00%			3148844	3	13.04%				560752	10	41.67%	
9264148	7	35.00%			560713	3	13.04%				560849	8	33.33%	
560999	6	30.00%			6	1	4.35%				8492512	8	33.33%	
11345232	5	25.00%									560532	7	29.17%	
12941730	4	20.00%									6993893	6	25.00%	
5726485	4	20.00%									6925826	4	16.67%	
6	4	20.00%									9106499	4	16.67%	
3175153	3	15.00%									13412010	3	12.50%	
4936828	3	15.00%									13412649	3	12.50%	
560464	3	15.00%									560427	1	4.17%	
10716070	2	10.00%									560504	1	4.17%	
1898401	1	5.00%												

Fig. 18. Degree of Resource Involvement in Process Stages for each Municipality

3.5.5 Analysis results

As the results of this analysis, we found that five activities of which the concept:name has the form of '99_NOCODE_##' are only operated by specific operators involved in various stages. (Cf. <Fig. 19>)

We called these activities as 'NOCODE Activity', including:

- cancel case (99_NOCODE_01)
- received OLO documents (99_NOCODE_02)
- process received advice (99_NOCODE_03)
- send letter receptive test not ok wait for report firedepartment (99_NOCODE_04)
- change authorized (99_NOCODE_05)

Although there were some exceptional cases, we understood them as unique characteristics of each municipality. For example, in case of the municipality 1, '267601' is the only top-level operator operating NOCODE Activity. Regarding the municipality 3, NOCODE Activity has never been found. And '560849' of municipality 4 was involved in the process stages by 65.38%, but didn't operate NOCODE Activity. Lastly, in case of municipality 5, we figured out that the operators numbered as '560530', '560752', '560532' operated NOCODE Activity even though they were involved in the process stages only as 50%. However, we found that they are the Resources moved from municipality 2 and municipality 4. (cf. paragraph 4.4) It seems that they had already been involved in various process stages in their each municipality. ['560530': Municipality 2 (20/26, 76.92%), '560532': Municipality 2 (21/26, 80.77%), '560752': Municipality 5 (23/26, 88.46%)]

Municipality 1 (20 process stages)			Municipality 2 (26 process stages)			Municipality 3 (23 process stages)			Municipality 4 (26 process stages)			Municipality 5 (24 process stages)		
Resource	# of process stages involved	%	Resource	# of process stages involved	%	Resource	# of process stages involved	%	Resource	# of process stages involved	%	Resource	# of process stages involved	%
2670601	19	95.00%	560458	24	92.31%	3442724	21	91.30%	1550894	23	88.46%	560604	23	95.83%
3273854	18	90.00%	560519	23	88.46%	560454	20	86.96%	560752	23	88.46%	560602	22	91.67%
560872	18	90.00%	4634935	21	80.77%	560749	20	86.96%	560821	23	88.46%	560429	21	87.50%
560462	17	85.00%	560521	21	80.77%	2013365	18	78.26%	560852	23	88.46%	560600	21	87.50%
560890	17	85.00%	560532	21	80.77%	560741	17	73.91%	560781	22	84.62%	560594	19	79.17%
560912	16	80.00%	560530	20	76.92%	560673	16	69.57%	560849	17	65.38%	560598	19	79.17%
560950	15	75.00%	20987361	8	30.77%	560696	16	69.57%	560812	11	42.31%	560608	18	75.00%
560894	14	70.00%	22445896	8	30.77%	5025869	10	43.48%	560431	6	23.08%	1254625	17	70.83%
11744364	11	55.00%	560598	8	30.77%	3122446	9	39.13%	560796	4	15.38%	560596	17	70.83%
560925	9	45.00%	560429	3	11.54%	560922	6	26.09%	6	2	7.69%	560530	10	41.67%
560589	8	40.00%	560528	3	11.54%	3069866	4	17.39%				560613	10	41.67%
560881	7	35.00%			3148844	3	13.04%				560752	10	41.67%	
9264148	7	35.00%			560713	3	13.04%				560849	8	33.33%	
560999	6	30.00%			6	1	4.35%				8492512	8	33.33%	
11345232	5	25.00%									560532	7	29.17%	
12941730	4	20.00%									6993893	6	25.00%	
5726485	4	20.00%									6925826	4	16.67%	
6	4	20.00%									9106499	4	16.67%	
3175153	3	15.00%									13412010	3	12.50%	
4936828	3	15.00%									13412649	3	12.50%	
560464	3	15.00%									560427	1	4.17%	
10716070	2	10.00%									560504	1	4.17%	
1898401	1	5.00%												

Fig. 19. Activity Operator Performing NOCODE Activity (red-colored)

Meanwhile, as seen in <Fig. 20>, there are some differences in every NOCODE Activity of each municipality.

Activity	concept:name	Municipality 1	Municipality 2	Municipality 3	Municipality 4	Municipality 5
cancel case	99_NOCODE_01					
received OLO documents	99_NOCODE_02					
process received advice	99_NOCODE_03					
send letter receptive test not ok wait for report firedepartment	99_NOCODE_04					
change authorized	99_NOCODE_05					

Fig. 20. Differences among the NOCODE Activities of each Municipality

3.6 Question 5: What are the differences in control flow between the municipalities?

3.6.1 Understanding the Question

Reference Process Model reflects rules or policies regulated in law or policies, which describes how to work effectively and efficiently. We decided to figure out the differences between reference processes model and workflows for each municipality.

3.6.2 Analysis

The followings are three points to analyze the differences between reference process model and workflows for each municipality:

- **Non-sequential Cases (Out-of-Order):** We figured out actual workflows by using given Case IDs and Timestamps. We ordered events by timestamps occurred at a same process stage such as main or service process. And we figured out some events which don't follow the order of reference process model. We decided to consider those events as out-of-order event. As seen in <Fig. 21>, the event which has '08_AWB45_20_2' as an action code occurs after the event which has an action code, '08_AWB45_025'. Based on our decision as above, we consider the case as out-of-order.
- **Cases with more than two same action codes (rework):** We decided to consider the case, which has more than two same action codes, as the rework case. For instance, the case identified by the Case ID '21381083' has two action_codes, '08_AWB45_005'. As our decision mentioned above, the case is a rework case (Cf. <Fig. 21>).
- **The combination of values of a municipality and a process stage without any case performed:** We listed municipalities and looked over the frequency of each process stage, which has no case but is in the Reference process model. The null value of action_code is recorded as '99_NOCODE_##'.

Case ID	Activity	Complete Timestamp	action_code	out_order	out_order (#)	rework	rework (#)
21381083	request complete	2014-03-25 8:00:00	08_AWB45_005	-	11	-	1
21381083	term for supplying missing data	2014-03-25 8:00:00	08_AWB45_010	-	11	-	1
21381083	create letter requesting missing data	2014-03-25 8:00:00	08_AWB45_020_0	-	11	-	1
21381083	retrieve missing data	2014-03-25 8:00:00	08_AWB45_020_1	-	11	-	1
21381083	phase asked additional information	2014-03-25 8:00:00	08_AWB45_025	-	11	-	1
21381083	enter senddate retrieve missing data	2014-03-25 21:59:57	08_AWB45_020_2	X	11	-	1
21381083	receive additional information	2014-05-01 7:00:00	08_AWB45_030	-	11	-	1
21381083	enter receipt of additional data	2014-05-01 7:00:00	08_AWB45_040	-	11	-	1
21381083	phase additional information received	2014-05-01 7:00:00	08_AWB45_045	-	11	-	1
21381083	send confirmation receipt additional data	2014-05-01 7:00:00	08_AWB45_051_0	-	11	-	1
21381083	request complete	2014-05-13 7:00:00	08_AWB45_005	X	11	X	1
21381083	keep permanently suspended	2014-05-13 7:00:00	08_AWB45_060	-	11	-	1

Fig. 21. Example of Analysis of Out-of-order and Rework Cases

3.6.3 Selecting of Target Data for Analysis

Incomplete cases can distort the analysis results related to Out-of-Order and rework events as mentioned above. Therefore, we selected only complete cases as the subject data to the analysis. And all events are considered in figuring out process stages without any case. The reason we decided to consider all events is when the incomplete cases are deleted the events related to the case are also deleted at the same time even though there could be some stages of the event.

3.6.4 Analysis results

3.6.4.1 Out-of-Order

<Fig. 22.> shows the summary of the analysis results of the Out-of-Order events. We found that most of the complete cases occurred without following the order. We also figured out that the average frequency of out-of-order events per one case in five municipalities is 13.46. In case of ‘municipality 2’, it has the highest number of occurrences and the highest rate (98.21%) of Out-of-Order cases for the overall cases.

	A: # of completed cases	B: total events	C: # of out_order cases	D: # of out_order events	C/A (%)	Average events per a completed case (B/A)	Average out_order per a completed case (D/A)	Average out_order per a out_order case (D/C)
Municipality 1	171	8,096	165	2,213	96.49%	47.35	12.94	13.41
Municipality 2	614	35,989	603	9,399	98.21%	58.61	15.31	15.59
Municipality 3	53	2,188	49	719	92.45%	41.28	13.57	14.67
Municipality 4	98	3,710	87	1,248	88.78%	37.86	12.73	14.34
Municipality 5	797	42,583	701	10,175	87.95%	53.43	12.77	14.51

Fig. 22. The Number of Out-of-Order of each Municipality

3.6.4.2 Rework

<Fig. 23> is the summary of the analysis of reworks. On average, there occurred 1.08 times of reworks per one case for five municipalities. Also, the highest rate of the reworks was from Municipality 5 as 48.06% and the lowest rate from Municipality 1

as 33.92%, for every completed case. The average number of reworks of every five municipality was 2.56. Additionally, in case of Municipality 2, there were 31 times of reworks only related to one case.

	A: # of completed cases	B: total events	C: # of rework cases	D: # of rework events	C/A (%)	Average events per a completed case (B/A)	Average rework per a completed case (D/A)	Average rework per a rework case (D/C)	Max (rework per a case)	Min (rework per a case)
Municipality 1	171	8,096	58	150	33.92%	47.35	0.88	2.59	17	0
Municipality 2	614	35,989	224	759	36.48%	58.61	1.24	3.39	31	0
Municipality 3	53	2,188	24	50	45.28%	41.28	0.94	2.08	8	0
Municipality 4	98	3,710	42	104	42.86%	37.86	1.06	2.48	8	0
Municipality 5	797	42,583	383	1,029	48.06%	53.43	1.29	2.69	26	0

Fig. 23. The Number of Reworks of each Municipality

3.6.4.3 Unperformed Process Stages and the Characteristics of its Frequency for each Municipality

<Fig. 24> refers to the rate of the occurrence of process stages for each municipality. We analyzed that the overall rate of process stages over all of the municipalities are similar to each other. And also we found that there are some unperformed process stages for each municipality (gray colored). Meanwhile, '01_HOOFD' and '08_AWB45' seem to be performed the most frequently over all of the five municipalities.

NO	process stage	M1		M2		M3		M4		M5	
		#	%	#	%	#	%	#	%	#	%
1	01_HOOFD	39,724	76.07%	32,018	72.19%	45,557	76.33%	34,523	73.00%	43,019	72.81%
2	01_BB	502	0.96%	1,905	4.29%	135	0.23%	1,604	3.39%	2,085	3.53%
3	01_OLO	-	0.00%	55	0.12%	10	0.02%	106	0.22%	12	0.02%
4	02_DRZ	832	1.59%	594	1.34%	884	1.48%	727	1.54%	769	1.30%
5	02_OLO	-	0.00%	1	0.00%	-	0.00%	-	0.00%	-	0.00%
6	03_GBH	169	0.32%	426	0.96%	909	1.52%	705	1.49%	713	1.21%
7	03_VD	49	0.09%	21	0.05%	27	0.05%	28	0.06%	10	0.02%
8	04_BPT	1,251	2.40%	1,525	3.44%	1,534	2.57%	1,289	2.73%	1,237	2.09%
9	05_EIND	865	1.66%	666	1.50%	585	0.98%	748	1.58%	1,037	1.76%
10	06_OLO	-	0.00%	1	0.00%	-	0.00%	1	0.00%	-	0.00%
11	06_VD	1,630	3.12%	945	2.13%	979	1.64%	772	1.63%	615	1.04%
12	07_OPS	172	0.33%	150	0.34%	343	0.57%	235	0.50%	296	0.50%
13	08_AWB45	3,325	6.37%	2,116	4.77%	3,517	5.89%	2,256	4.77%	4,124	6.98%
14	08_AWB45_WAW	2	0.00%	-	0.00%	-	0.00%	-	0.00%	1	0.00%
15	08_OLO	-	0.00%	5	0.01%	1	0.00%	10	0.02%	1	0.00%
16	09_AH_I	957	1.83%	709	1.60%	1,248	2.09%	883	1.87%	991	1.68%
17	09_AWB45	154	0.29%	102	0.23%	125	0.21%	146	0.31%	179	0.30%
18	10_OLO	-	0.00%	4	0.01%	2	0.00%	5	0.01%	-	0.00%
19	10_UOV	878	1.68%	1,454	3.28%	926	1.55%	876	1.85%	861	1.46%
20	11_AH_II	1,045	2.00%	687	1.55%	1,364	2.29%	786	1.66%	799	1.35%
21	11_OLO	-	0.00%	-	0.00%	-	0.00%	2	0.00%	1	0.00%
22	12_AP	39	0.07%	25	0.06%	646	1.08%	6	0.01%	121	0.20%
23	12_AP_UOV	15	0.03%	12	0.03%	22	0.04%	15	0.03%	28	0.05%
24	13_CRD	386	0.74%	294	0.66%	501	0.84%	435	0.92%	419	0.71%
25	14_VRU	169	0.32%	115	0.26%	242	0.41%	256	0.54%	779	1.32%
26	15_NGV	50	0.10%	64	0.14%	119	0.20%	45	0.10%	191	0.32%
27	16_LGSD	-	0.00%	182	0.96%	1	0.01%	332	1.49%	317	1.25%
28	16_LGSV	2	0.00%	243	0.55%	4	0.01%	372	0.79%	421	0.71%
29	99_NOCODE	1	0.00%	35	0.08%	-	0.00%	130	0.27%	57	0.10%
	total	52,217	100.00%	44,354	100.00%	59,681	100.00%	47,293	100.00%	59,083	100.00%

Fig. 24. Relation to Process Stage, Action_code, and Activity of Five Municipalities

3.7 Question 6: Some of the procedures will be outsourced from 2018, i.e. they will be removed from the process and the applicant needs to have these activities performed by an external party before submitting the application. What will be the effect of this on the organizational structures in the five municipalities?

3.7.1 Understanding the Question

Outsourcing, which could be one of the methods of cost reduction, is a way to control some support activities, excepting the primary activities in an organization, through the other parties. Through this, we such organization could increase its concentration on tasks and professionalism by assigning its primary resources proper tasks. Generally, such methods outsource some high-matured process which is easy to control. Based on the process stages dealt with in the paragraph 4.1, we intend to figure out the which procedures are highly matured and the effect after outsourcing them, the high-matured procedures.

3.7.2 Selecting the suitable Municipality for the Question

We firstly needed to select some proper procedures to be target of outsourcing. Here, regarding of the main procedures (01_HOOFD), we decided not to include the process in this analysis because it was impossible to remove from the processes. We also selected some procedures which have high frequencies of performance to get greater effect. Finally, as the subjects to this analysis, we found four sub-procedures which have been operated more than 2,000 times total in the five municipalities: '08_AWB45 (5,429 events)', '01_BB (4,081 events)', '04_BPT (2,238 events)', '10_UOV (2,135 events)' (Cf. paragraph 3.6.4)

3.7.3 Analysis

We planned to evaluate the maturity of each procedure with (1) how well it is managed (Whether the procedure occurs in order corresponding the reference process model and doesn't accompany any rework). (2) How easy to manage/control it (The level of its standardization).

3.7.3.1 How Well the Procedure Is Managed

We, based on the paragraph 3.6.2, planned to find the number of performances and reworks for each procedure which corresponds to the order of the process, by comparison with the reference process model.

3.7.3.2 How Easy to Manage the Procedure

We tried to evaluate the level of the standardization of relevant procedure by understanding the method and the rate of the procedures operated for each procedure.

3.7.4 Analysis results

3.7.4.1 How Well the Procedure Is Managed

<Fig. 25> is the summary of the results of analysis referring to the paragraph 3.7.3.1. '10_UOV' shows the highest rate as out-of-order by 35.08%, followed by '08_AWB45' as 25.36% and '01_BB' as 17.52%. And '08_AWB45' topped the list of the numbers of reworks with 4.18% while the other three procedures were less than 1%. Based on the analysis results as mentioned above, we found that the procedure '04_BPT' is currently well managed.

Process	A: process total (#)	B: out of order		=B/A (%)	Process	A: process total (#)	B: rework		=B/A (%)
		status	#				status	#	
08_AWB45	5,429	-	4,052	74.64%	08_AWB45	5,429	-	5,202	95.82%
		x	1,377	25.36%			x	227	4.18%
01_BB	4,081	-	3,366	82.48%	01_BB	4,081	-	4,056	99.39%
		x	715	17.52%			x	25	0.61%
04_BPT	2,238	-	2,234	99.82%	04_BPT	2,238	-	2,232	99.73%
		x	4	0.18%			x	6	0.27%
10_UOV	2,135	-	1,386	64.92%	10_UOV	2,135	-	2,130	99.77%
		x	749	35.08%			x	5	0.23%
Total	13,883		13,883	100.00%	Total	13,883		13,883	100.00%

Fig. 25. The Number of Out-of-Order (left) and Rework (right) for Each Process.

3.7.4.2 How Easy to Manage the Procedure

We drew top six (6) variants for each procedure selected in the paragraph 3.7.2. As seen in <Fig. 26-29.>, among the six variants, '04_BPT' shows the highest accumulative rate with 95.03%, followed by '01_BB' as 85.83% and '08_AW045' as 69.65%. '10_UOV' has dramatically lower rate with 16.35%. So we concluded that '04_BPT' and '01_BB' are the procedures which are easy to manage as they are standardized.

No	Variant (156 in Total)	Case				Activity
		#	%	accumulation (#)	accumulation (%)	
1	A	564	55.40%	564	55.40%	A request complete
2	B→C→D→E→F→G	44	4.32%	608	59.72%	B enter senddate retrieve missing data
3	A→C→D→B→H→F→E→I→J→A	38	3.73%	646	63.46%	C term for supplying missing data
4	A→C→K→D→H→B→A→F→E→I→J	22	2.16%	668	65.62%	D retrieve missing data
5	C→D→B→F→E	22	2.16%	690	67.78%	E enter receipt of additional data
6	B→A→C→D→H→F→E→I→J→A	19	1.87%	709	69.65%	F receive additional information
						G MER present in supplement
						H phase asked additional information
						I phase additional information received
						J send confirmation receipt additional data
						K create letter requesting missing data

Fig. 26. Top Six Variants of Procedure '08_AW045'

No	Variant (32 in Total)	Case				Activity
		#	%	accumulation (#)	accumulation (%)	
1	A → B	407	27.99%	407	27.99%	A objection lodged against decision
2	A → C → B	390	26.82%	797	54.81%	B set phase: phase permitting irrevocable
3	A → B → C	183	12.59%	980	67.40%	C phase decision irrevocable
4	A → D → B	137	9.42%	1,117	76.82%	D set phase decision revoked
5	B → A	79	5.43%	1,196	82.26%	
6	B → A → C	52	3.58%	1,248	85.83%	

Fig. 27. Top Six Variants of Procedure '01_BB'

No	Variant (24 in Total)	Case				Activity
		#	%	accumulation (#)	accumulation (%)	
1	A	859	76.22%	859	76.22%	A regular procedure without MER
2	A → B → C → D → E → F	99	8.78%	958	85.00%	B MER present
3	B → C → F	39	3.46%	997	88.46%	C activities regular procedure
4	G → B → C → F	32	2.84%	1,029	91.30%	D start WABOprocedure
5	A → B → C → F	32	2.84%	1,061	94.14%	E read field phased application
6	B → C → F → H → I	10	0.89%	1,071	95.03%	F phased application
						G no permit required for application
						H first or second phase
						I term 14 or 26 weeks

Fig. 28. Top Six Variants of Procedure '04_BPT'

No	Variant (24 in Total)	Case			
		#	%	accumulation (#)	accumulation (%)
1	A → B → C → D → E → F → G → H	7	3.37%	7	2.27%
2	C → D → E → F → G → B → A → H	6	2.88%	13	6.25%
3	C → I → G → J → B → A → H	6	2.88%	19	9.13%
4	C	5	2.40%	24	11.54%
5	C → I → G → B → J → A → H	5	2.40%	29	13.94%
6	A → K → B → C → E → F → G → H → I	5	2.40%	34	16.35%

Activity	
A	date draft decision for inspection
B	entering senddate draft decision environmental permit
C	open draft decision for inspection
D	term perspective on draft decision
E	publish official draft decision
F	phase draft decision made
G	send design decision to stakeholders
H	perspective on draft decision environmental permit received
I	create publication document
J	enter senddate draft decision
K	publication date draft decision
L	final VVGB applications

Fig. 29. Top Six Variants of Procedure '10_UOV'

3.7.4.3 Effects of Outsourcing Procedures on Each Municipality

As the results of the analysis the paragraph 3.7.4.1-2, only '04_BPT' is the procedure which is highly matured and satisfy both two conditions. So we tried to find how outsourcing the procedure affect the process for each municipality. There are some information about events of the procedure occurred in each municipality and its rate in <Fig. 30>. '05_BPT' has 1.37% for five municipalities on average. If the procedure is outsourced, the process will be also removed and relevant Resources who have taken charge of these procedure can be moved to other duties. Therefore, it is expected that the relevant resources can be assigned to another procedure or activities for the previous rate the procedure have occupied were reduced.

Process	Municipality 1			Municipality 2			Municipality 3			Municipality 4			Municipality 5		
	#	%	process Total (#)	#	%	process Total (#)	#	%	process Total (#)	#	%	process Total (#)	#	%	process Total (#)
04_BPT	199	1.23%	16,192	1,116	1.54%	72,369	71	1.62%	4,376	117	1.58%	7,414	735	0.86%	85,142

Fig. 30. The Rate of the Procedures Targeted for Outsourcing for Each Municipality

However, there is a limitation in this analysis above that we didn't consider there could more procedures or activities be generated through the outsourcing.

4. Conclusions

We were provided some actual log data of the building permit application process of five Dutch municipalities through the BPI Challenge 2015. The five logs contain 1199/832/1409/1053/1156 cases and 52217/44354/59681/47293/59083 events respectively. We were also provided six questions related to the building permit application process. And we arranged one more separate question '0' which could be helpful for us to understand the building permit application process and answer the provided questions. We have got answers to the seven questions in various ways as below:

0. Definition of Completed Case, Process Stage, and Reference Process Model

We defined completed cases for our further analysis regarding the six given questions. The processes, composing the building permit application process, were subdivided as we utilized 'action_code' column; then we

could draw a reference process model based on the subdivided processes. (Cf. Figures in Analysis Results of Question 0)

1. Improvements of Organizational Structure

We suggested a segregation of duties as one possible means to improve the organizational structure. We also found, as a conclusion of the analysis, that the segregation of duties was poorly achieved in all of the five municipalities (Cf. Figures in Analysis Results of Question 1).

2. Differences among Process Throughput Times

We identified parameter causing longer throughput times of processes and explained the reason of the differences the municipalities. Following our analysis, we found that the process throughput times are delayed when the case procedure performance ((case) caseProcedure Column) is extensive (Uitgebreid) (Cf. Figures in Analysis Results of Question 2).

3. Employee Transferences and Changes in the Process

There are four employees who were moved from Municipality 2 and 4 to Municipality 5. The movement caused some various changes in the process of municipality 5. We found some critical changes in the beginning and the latter parts of the target process, by comparing the task performance methods before and after the transference in municipality 5 (Cf. Figures in Analysis Results of Question 3).

4. Characteristics of Resources Operating Various Stages of Process

The process stages for each municipality are different and so are the involvements of each process operator. We figured out some employees involved in various stages (more than 80% of involvements) and also five NOCODE Activities which are only performed by them as their unique task (Cf. Figures in Analysis Results of Question 4).

5. Comparison Reference Process Model and Workflow of Each Municipality

We utilized the reference process model drawn from Question 0 and compared each control flow of Municipalities. In detail, we found some out-of-order processes, reworked processes, and some process stages which are located in the reference process model, but never performed, through the comparison of the workflows of each municipality and the reference process model (Cf. Figures in Analysis Results of Question 5).

6. The Effect of Outsourcing

We firstly selected some high-matured process as the subject to the outsourcing. And we analyzed the effect of outsourcing of the targeted procedures for each municipality by focusing and examining the importance of the procedures in each municipality (Cf. Figures in Analysis Results of Question 6).

When applying our analysis results to any actual business level, there could be a limitation because we analyzed the questions above without any feedback of the five municipalities in Dutch. So there will be more actual and useful insight with a combination of our analysis results and the domain knowledge of any person concerned. Finally, we strongly recommend that more e-government managers pay

attention to process mining and other analytical techniques for better decision making for e-government initiatives in a big data world [6,7,8,9].

5. Recommendation

5.1 Applying New Business Rules and Monitoring

We found that segregation of duties had not been performed well in the overall five municipalities. There could be a higher possibility of intra-organizational corruption, without proper segregation of duties. We, therefore, recommend to build some business rules related to the segregation of duties and continuously monitor them.

5.2 A Thorough Management of the Critical NOCODE Activity

All of the operators, who performed NOCODE activity, were involved in various procedures and every activity belonging to the NOCODE Activity were meaningfully critical activities. There seems to be a high possibility that relevant managers operate the NOCODE Activity. Therefore, there should be some clearly designated KPIs (Key Performance Indicators) about NOCODE Activity and continuous achievement management.

5.3 Applying Different Management Methods for the Each Different Throughput Time

We found the parameter (caseProcedure) which affect the throughput time of each case. So we could anticipate through this parameter that there will be some longer throughput times of cases. We could reduce the throughput time by anticipating processes instances possible to have longer throughput times and intensively managing them.

5.1 Building New Reference Process Model and Employee Training

We figured out that process workflows could be changed through any employee movements. It means that task performance methods of each employee affect processes more largely than applying Reference Process Model. We also found there are many cases of violation of the Reference Process Model (Out-of-order, rework, unperformed). And it refers that either the Reference Process Model is so inefficient that it doesn't reflect any actual business methods or the employees does not follow the Reference Process Model for performing their duties. In case of the former, another new Reference Process Model which reflects the actual business should be organized again; in case of the latter, the employee needs to be educated to follow the Reference Process Model through enhanced employee training.

Acknowledgements

The authors are grateful to Dr. Anne Rozinat and Dr. Christian Gunther (Fluxicon) for providing them with an evaluation version of Disco and an accompanying copy of the BPIC 2015 datasets. They also thank Chang Jae Kang (PMIG: Process Mining & Intelligence Group) for our thoughtful guidance and Sara Lee for her translation help.

References

1. Grönlund, Åke, ed. *Electronic Government: Design, Applications and Management: Design, Applications and Management*. IGI Global, 2001.
2. W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, H.M.V. Verbeek, *Business process mining: An industrial application*, *Information Systems* 32 (5), 2007, pp. 713-732.
3. W.M.P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer, 2011.
4. Hollenbeck, John R., Barry Gerhart, and Patrick M. Wright. *Fundamentals of human resource management*. Vol. 2. McGraw-Hill, 2004.
5. Rummier, GA., Brache, AP.: *Improving performance: How to manage the white space on the organization chart*, 2012.
6. A. McAfee, E. Brynjolfsson, *Big data: The management revolution*, *Harvard Business Review* 90 (10), 2012, pp. 60-69.
7. Á. Rebugue, D.R. Ferreira, *Business process analysis in healthcare environments: A methodology based on process mining*, *Information Systems* 37 (2), 2012, pp. 99-116.
8. A. Rozinat, W.M.P. van der Aalst, *Conformance checking of processes based on monitoring real behavior*, *Information Systems* 33 (1), 2008, pp. 64-95.
9. S. Suriadi, M.T. Wynn, C. Ouyang, A.H.M.t. Hofstede, N.J.v. Dijk, *Understanding process behaviours in a large insurance company in Australia: A case study*, *Advanced Information Systems Engineering Lecture Notes in Computer Science* 7908, 2013, pp. 449-464.