# Rabobank: Incident and change process analysis

Michael Arias [1], Mauricio Arriagada[1], Eric Rojas[1], Cecilia Sant-Pierre[1], Marcos Sepúlveda[1]

[1] Pontificia Universidad Católica de Chile,
Av. Vicuña Mackenna 4860, Macul,
Santiago, Chile
{michael.arias, marriagb, eurojas, csaintpierre}@uc.cl, marcos@ing.puc.cl

**Abstract.** The purpose of this report is to present the results of a work performed as part of the Fourth International Business Process Intelligence Challenge (BPIC´14). This challenge provides a real life event log from Rabobank Netherlands Group ICT, a log that contains information related with service desk processes, including interaction, incident and change management. We show the analysis performed applying different tools, including a prediction analysis, impact patterns analysis, change process review and the use of process mining techniques to analyze process characteristics and team's interaction. The results generated can be useful for Rabobank, providing them more knowledge about the incident and change management process, and also, bringing some insights that can help change implementation teams in their tasks related to improve their standard operation procedures.

**Keywords**: Incident management, change management, service desk, impact patterns, process mining, business process, organizational mining, Rabobank Group.

## 1 Introduction

As part of business process improvements, many companies have to apply changes in their processes in order to succeed. As an example, some IT companies has introduced ITIL (Information Technology Infrastructure Library) processes to align IT services on business, such is the case of Rabobank Group ICT.

Expecting to implement planned changes, Rabobank uses Change-process from ITIL-processes. In this implementation, the company wants to observe for fact-based insight into the impact of changes at the Service Desk and IT Operations.

In this paper, we analyze the data presented by Rabobank Group ICT. We present our founding based on the questions presented in the Business Process Intelligence Challenge 2014: Identification of Impact-patterns, Parameters for every Impact-pattern, Change in average steps to resolution and creativity challenge. We applied Process Mining techniques [1] to analyze the data using ProM [2] and Disco [3] tools. In addition we applied Data Mining tools [4] to process and estimate changes in advance using WEKA software [5]. In the following sections we describe the data, and presents a work findings.

## 2 Available data set

The information available to respond to the Rabobank interest were the data related to the ITIL process implemented in the Bank. The process (Figure 1) starts with a customer (internal client) that contact by call or mail to the Service Desk to report disruption on some ICT-service. A Service Desk Agent receives the customer contact and registers the information about the service disrupted and the configuration Item affected. This information is recorded in an event log including user, time-stamp and agent code. The calls that are not solved directly are registered as an incident and assigned the task to a expert team to solve it. The SD could receive several calls reporting the same disruption that will be associated to a same incident case. A particular disruption could reoccur more often than usual, so a Problem Investigation is started, if is the case, a Request for Change will be created and the change case will be assigned to an expert.

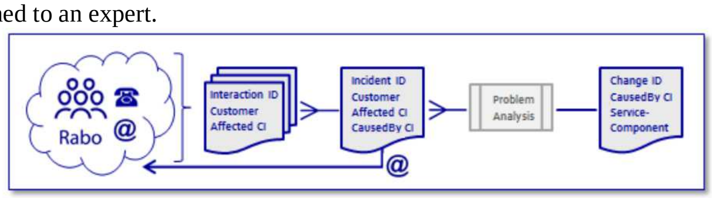



**Fig. 1.** Data relationship.

The available data consisted in the event logs from the Service Management Tool used by Rabobank to manage the ITIL processes. That was composed by four tables with the information of interactions records, incidents records, incident activities and changes records, performed since 2012 until march 2014, been the most significant activity between August 1$^{st}$ 2013 and march 31st 2014. There also some isolated data before 2012 and after April 2014, but were considered outlier cases

Interaction table has 102.461 records, each one corresponding to an interaction. Every record contain an id code, information about the interaction category (incident or request for information), whether it was solve in the first call or not, impact, urgency and priority of the call, among other information.

Incident table has 46.606 records, each one corresponding to an incident case. Every record has an id code, the id code to the related interaction, if is just one, or a #multivalue indicating there was more than one interaction related. The record also has information about the Configuration Item affected, the WBS affected and which CI and WBS caused the disruption.

Incident activities table has 343.121 records, linked to an incident id, with the activities performed on each case. In order to have a better understanding of the process, the interactions and incident tables were merged and used to our analysis.

Finally, the changes table contain the records of the activities performed on each change case, including information about configuration item affected, Service component affected, change type and risk assessment, among others. The records also include information of the following dates: Planned Start, Planned End, Scheduled

Downtime Start, Scheduled Downtime End, Actual Start, Actual End, Requested End Date, Change record Open Time, Change record Close Time.

## 3 Identification of impact patterns

Understanding the interest of Rabobank of having enough information to plan the Service Desk offering in advance, we explored the data searching for relationships between the amount of incidents and the change cases opened and change cases closed in a period of time. In our analysis, we used a week as time unit because it is enough time to make changes to a staff planning, and it also allows us to make better predictions.

As a first approach to understand the relationship between the changes implementation and the service desk workload, we compared the number of opened/closed change cases by week with the workload of the Service Desk measured as the number of incidents and requests for information (RFI). We identified a correlation between both variables, observing a demand increase/decrease in the SD workload when the change activities increased/decreased. We can observe, for example, a period, which is consistent with the Christmas holidays, when the number of changes decreases and, at the same time, the number of incidents also decreases.

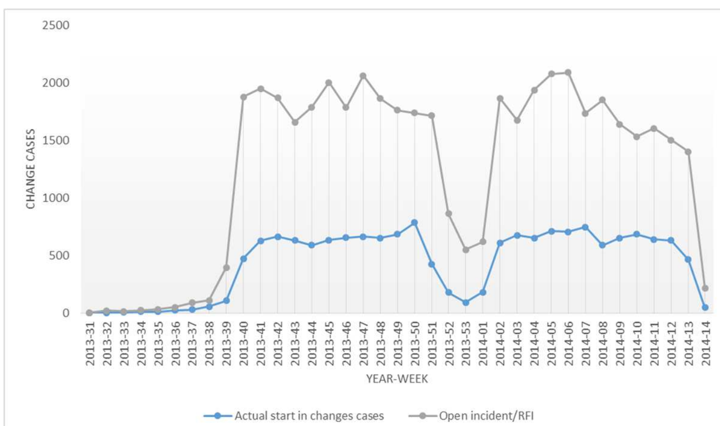



**Fig. 2.** Comparison between Service Desk workload and the number of opened/closed change cases, by week.

Based on the correlation detected, we created a prediction model for the number of incident cases of a week, considering as variables the number of change cases of the same period (supposing the change cases opened are predictable a week before), but also considering the SD incidents and change cases of the previous two weeks.

$$I_t = f(I_{t-2}, I_{t-1}, C_{t-2}, C_{t-1}, C_t),$$

where $I_t$ is the number of opened incidents in week t, and $C_t$ is the number of change cases opened in week t. We tried different supervised learning tools to obtain $I_t$, such as decision trees and neural networks (NN), obtaining the best results with a multilayer perceptron classifier with one hidden layer. The model obtained for 35 weeks has a correlation coefficient of 0.9825 and a relative absolute error of 16.92%. Details of the model are shown in Figure 3.

```
=== Classifier model ===

MLPRegressor with ridge value 0.01 and 2 hidden units

Output unit weight for hidden unit 0: 0.6688768674740564
Hidden unit weights:
-0.5335413081109858 T
-1.3193030695184804 changes (T)
-0.6959543778185515 Changes (T-1)
-0.5720355158281024 Incident (T-1)
0.07509303106339088 Changes (T-2)
-0.06594498871117522 Incident (T-2)
Hidden unit bias: 0.3097032540578573

Output unit weight for hidden unit 1: -3.5460301028697527
Hidden unit weights:
0.2962776311738805 T
-1.7079871084686875 changes (T)
-0.15546918648135305 Changes (T-1)
-0.793376613992151 Incident (T-1)
0.7294845275543771 Changes (T-2)
-0.5174566912036446 Incident (T-2)
Hidden unit bias: -1.957058808746062
```

**Fig. 3.** Output of the MLP regression obtained in the software WEKA.

The prediction model obtained is a good predictor of the next week workload, as shown in Fig. , but it might not be a simple task to know in advance the number of change cases scheduled for the next week, which is an input of this model, so as to make a good prediction. To avoid this limitation, we used the same data to create a model using only past information. Results are shown in Figure 4. The second model has a correlation coefficient of 0.9744 and a relative absolute error of 19.267%.

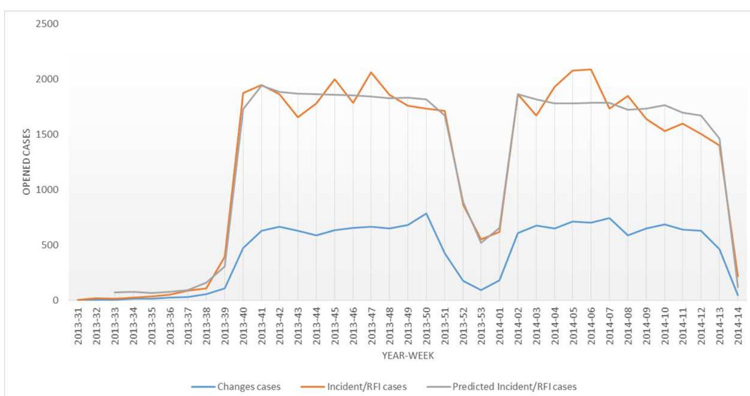



**Fig. 4.** SD workload predicted in comparison with the real data.

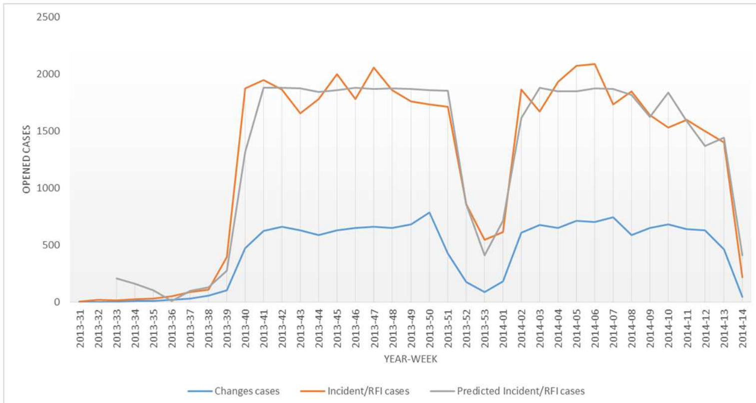



**Fig. 5.** SD workload predicted using only past data in comparison with the real data.

```
=== Classifier model (full training set) ===

MLPRegressor with ridge value 0.01 and 1 hidden units (useCGD=false)

Output unit weight for hidden unit 0: -4.673160449897752

Hidden unit weights:

0.7478684827533448 T
-1.4964597480335107 Changes (T-1)
-4.645221428518859 Incident (T-1)
4.923370012268863 Changes (T-2)
-1.2932242026396314 Incident (T-2)

Hidden unit bias: -3.7507593859854325

Output unit bias: 0.761322527397926
```

**Fig. 6.** Output of the second MLP regression model obtained in the software WEKA

Finally, we can observe in Figure 7 that the prediction plus/minus the absolute error generated by the model provides a very accurate range where the real value is contained.

With the results obtained from both models we can predict how the number of incidents will increase/decrease in general terms, i.e., the service desk demand, but this analysis does not consider the nature of the changes and/or the incidents.

In order to link the changes data with the number of reported incidents, we considered which configuration item type was impacted (CI Type Aff) in each case in the change log and how many incidents were reported for this particular case. We grouped this data by week, counting the CI Types affected and adding the amount of incident reported. The date considered to construct the table was the “Change record close time”, so as we could obtain the relationship between closed cases and related incident.

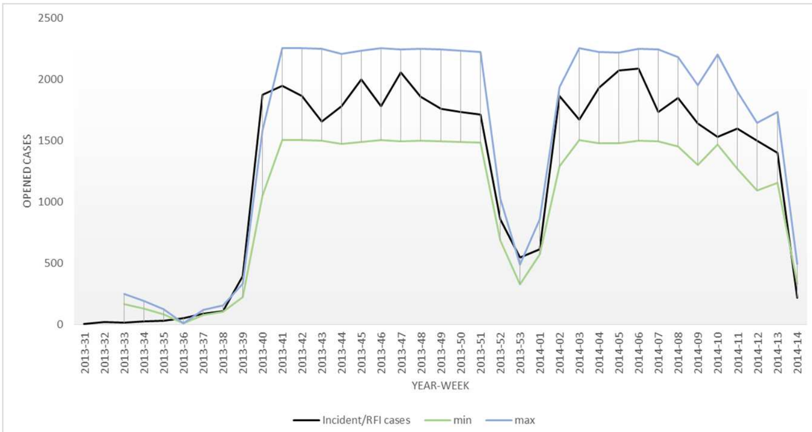



**Fig. 7.** Center line shows the real SD workload, while the line below and above present the range of the prediction ± absolute error.

In order to link the changes data with the number of reported incidents, we considered which configuration item type was impacted (CI Type Aff) in each case in the change log and how many incidents were reported for this particular case. We grouped this data by week, counting the CI Types affected and adding the amount of incident reported. The date considered to construct the table was the “Change record close time”, so as we could obtain the relationship between closed cases and related incident.

Based on this data, we made a prediction model that related the number of cases that impact every configuration item type (CIType) with the number of reported incidents, grouped by weeks. We used a multilayer perceptron classifier with one hidden layer to create the model, and the results had a correlation coefficient of 0.9981 and an absolute error of 6.3801%. The model is presented in Figure 8 (a).

Furthermore, using the same data, we created a model to predict the next week number of incidents related to the changes registered and CI Type impacted, and the model was also accurate in predicting the results. The model obtained is shown in Figure 8 (b), and a graphic that shows the real and predicted number of incidents related to the changes cases are shown in Figure 9.

```
=== Classifier model ===

MLPRegressor with ridge value 0.01 and 2
hidden units
Output unit weight for hidden unit 0:
2.154321965642582
Hidden unit weights:
0.5988457200066349 T
0.4865488268241506 displaydevice
-0.5373334416288943 Phone
-0.1542922071457798 networkcomponents
0.9604661593597394 officeelectronics
0.4549540911057183 applicationcomponent
0.823314239239705 database
-0.574768259158646 storage
-0.284199940351281 computer
1.0422377281027466 application
0.7956685470541865 software
-0.5277020206319407 no type
1.1669353391837611 subapplication
1.0998486519396427 hardware

Hidden unit bias: -3.440804425083141
Output unit weight for hidden unit 1:
2.779449815687286
Hidden unit weights:
1.4976867270776748 T
0.40792337178516846 displaydevice
0.903516900784622 Phone
0.24238451216250617 networkcomponents
0.5460731831723057 officeelectronics
-0.7014140116876212 applicationcomponent
-0.36196742344132476 database
1.995609684432313 storage
-0.436042375376004 computer
1.0834486689954592 application
-0.6132683974526137 software
0.6846857367104329 no type
-0.6417069269949371 subapplication
0.6481252495487153 hardware

Hidden unit bias: -2.269215602941567
Output unit bias: -0.9976217598938246
```

(a)

```
=== Classifier model ===

MLPRegressor with ridge value 0.01 and 2 hidden
units (useCGD=false)
Output unit weight for hidden unit 0:
3.118139062618232
Hidden unit weights:
0.2028076450399164 T
-0.13292046269679716 displaydevice
-0.21348745126341018 Phone
-0.2059648787599394 networkcomponents
-0.10299180973056202 officeelectronics
1.2876278493987654 applicationcomponent
0.11841725112013356 database
0.24489025221253 storage
0.6173186966632981 computer
-1.542388074534464 application
1.491139004568813 software
-1.0760732105781394 no type
-1.672704870811626 subapplication
2.814418269288297 hardware

Hidden unit bias: -2.41805963601873
Output unit weight for hidden unit 1: -
3.852205418857143
Hidden unit weights:
0.1867966057601190 T
-0.0017317226923799901 displaydevice
-1.888582915638658 Phone
-0.6636760602415008 networkcomponents
0.2714073516056565 officeelectronics
2.5050274802541157 applicationcomponent
0.4568989981256474 database
0.842993242960647 storage
-2.333367380681127 computer
-1.5381912815737937 application
2.339555687634827 software
-1.0358319426060738 no type
-1.788056092788681 subapplication
2.367575388716674 hardware

Hidden unit bias: 1.4993851142959937
Output unit bias: 1.7801030183978799
```

(b)

**Fig. 8. (a)** Output of the MLP regression obtained in the software WEKA to predict the number of incidents based on the CI Type impacted by the changes. **(b)** Output of the MLP regression obtained in the software WEKA to predict the number of incidents based on the CI Type impacted by the changes considering only past data.

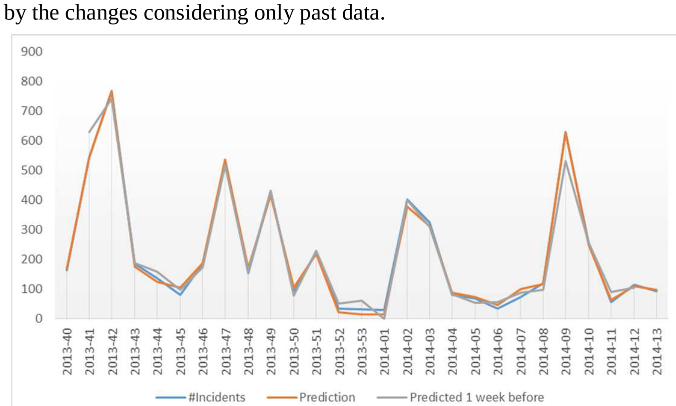



**Fig. 9.** Comparison between the real number of incidents related to change cases and the prediction obtained based in the CI Types affected by the changes and the same prediction performed one week in advance.

## 3 Parameters for every impact-pattern

At this point, we construct a model to have a very accurate prediction of the Service Desk workload in terms of quantity of incident reported, but we do not know how is impacted the work of the SD in terms of activities, duration or complexity of the cases. To clarify that, we perform an analysis of the interaction log using process mining tools.

We start our analysis preprocessing the information in the incident log. The result of this preprocessing was a set of 45.933 cases of incidents, and 456.622 events. This set was analyzed in the software Disco, to understand the process patterns that the SD is carrying out and link it with the workload predicted. The process model obtained from the preprocessed log, using with the Disco tool was a spaghetti-like model, as is shown in Figure 10.

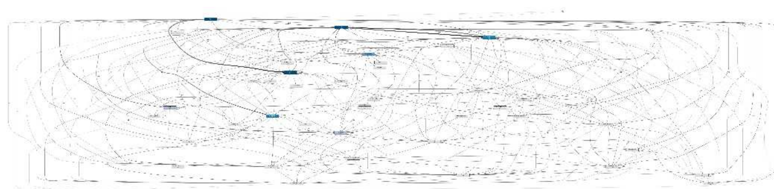


**Fig. 10.** Process model obtained using the raw incident log. There can be seen that the model is unreadable, so some segmentation is needed to understand the process.

The event log has 39 different activities, with a mean duration of 4.3 days. The most common activities are 7, which accumulate 81.7% of the events (Table 1). Most of the cases start with the Open activity, but few start with Closed, that was filtered considering there are not complete cases. The same analysis was performed with the endpoint, filtering the cases ended with the activity End and Caused by CI, the last one, because that represents a 33% of the cases.

**Table 1.** Most frequent activities in the preprocessed incident log.

| *Activity* | *Frequency* | *Relative frequency* |
|---|---|---|
| Assignment | 86574 | 18.96% |
| Operator Update | 54915 | 12.03% |
| Reassignment | 50728 | 11.11% |
| Status Change | 49775 | 10.90% |
| Closed | 49754 | 10.90% |
| Open | 45930 | 10.06% |
| Update | 35202 | 7.71% |

About the duration of the cases, a 97% of those finish before 30 days, so we filter the log discarding the long duration cases, trying to model the mainstream behavior. After those filters, the event log had 41.422 cases with almost 390 thousand events.

We noticed there are two important categories of cases in the dataset, those where the case start as a request for information (18.9%), and the ones which start as incident report (81.1%). There are also a few cases categorized as Compliant, but the category is not relevant for de analysis. The incidents data set has also an interesting categorization by the closure code, there could be identified an 82.6% of the cases where the closure code represent the area affected by the incident (hardware, software, other), and a 17.4% of the cases where the closure code indicates that there was not an incident (no error - works as designed, question, user manual not used). Table 2 shows the detail of the amount in each category.

**Table 2.** Division by closure code of the incident dataset.

| *Closure code* | *Incident* | *Request For Information* | *Total* |
|---|---|---|---|
| **Data** | 1547 | 659 | 2206 |
| **Hardware** | 2997 | | 2997 |
| **Software** | 12343 | 636 | 12979 |
| **Other** | 11043 | 5309 | 16352 |
| **Unknown** | 1368 | 206 | 1574 |
| **Referred** | 135 | 21 | 156 |
| **Operator error** | 1406 | 129 | 1535 |
| **No error - works as designed** | 2486 | 1024 | 3510 |
| **User error** | 3112 | 441 | 3553 |
| **User manual not used** | 664 | 101 | 765 |
| **Questions** | | 131 | 131 |
| **Inquiry** | | 161 | 161 |
| **Total** | **37101** | **8818** | **45919** |

In a first sight we expect that the cases started as request for information where related with closure codes as questions or user manual not used, but there is not a direct relation. This defines four quadrants we will to analyze as is shown in Figure 11.

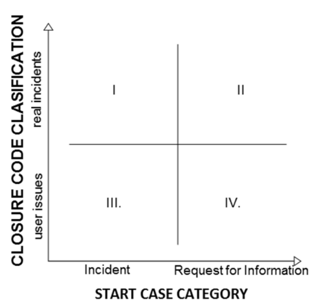



**Fig. 11.** Classification of the Service Desk cases by start category and closure code.

Every quadrant should be analyzed separately, since the clients and SD behavior is different in each case, and the relation between every case with the changes performed would be different.

**Quadrant I (67.4% of the cases)**

There is the biggest group and the correlation between both, changes cases opened and Service Desk workload is the same viewed in the complete dataset. In Figure 12 there is shown the comparison between both variables.

The cases in this quadrant are those cases where the user detects a wrong operation of a component (application, hardware, among others). The average duration of these cases is 66.8 hours.

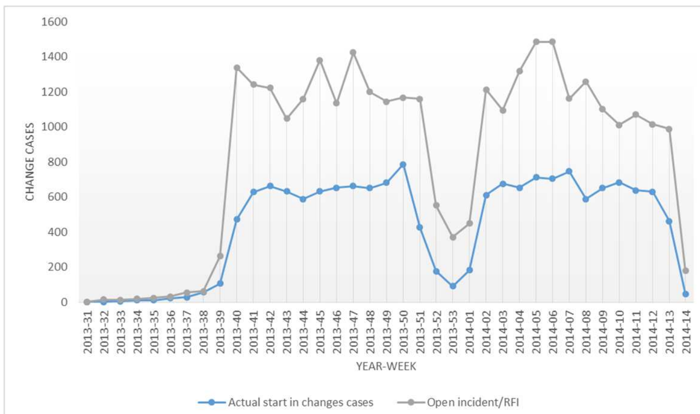



**Fig. 12.** Workload comparison between Service Desk and Changes Teams, expressed in cases opened every week, filtered by category as {Incident} and closure code as {Data, Hardware, Software, Other, Unknown, Referred, Operator error}

In this case, the process map is also a spaghetti-like model. We use the Flexible heuristic miner algorithm to discover the process, but the result obtained is highly complex with several parallel paths. This log segment has 27.457 cases and 13.995 variants, that shows there are multiples ways to perform the process. In this quadrant, there are 38 of the 39 total activities, Fig.  shows a best understandable process map obtained with the software Disco.

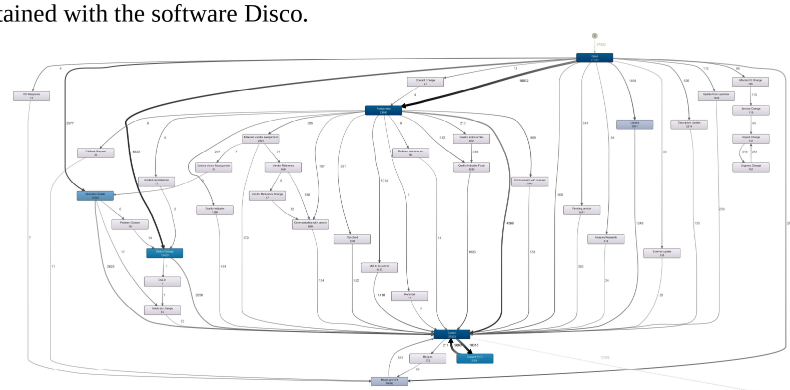


**Fig. 13.** Process map of the first quadrant obtained using the tool Disco.

In the process flow is interesting to note that in 2.746 cases, the first activity after open the case is reassignment, we consider this behavior is a wrong execution of the process. This activity is performed in 10.677 cases, once in every case.

Several activities in the process are related to update information of the case. Not considering that activities left us just 26 activities performed and the same amount of cases. We filtered also the activities with less than a 0.1% of occurrence which was 12 activities. The resultant model is shown in Figure 14 and the activities performed with its frequency in the Table 3.

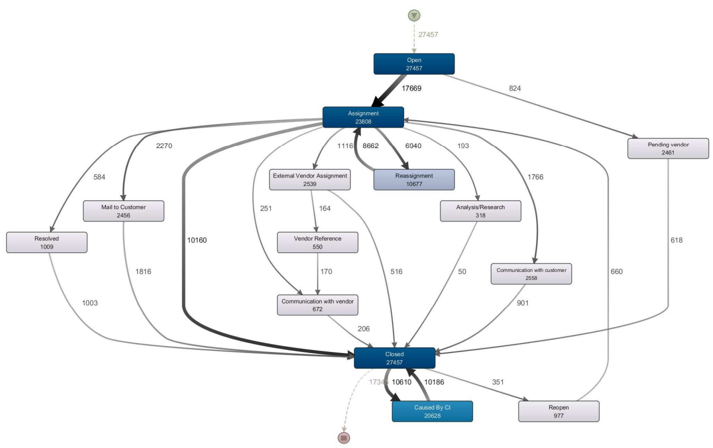



**Fig. 14**. Process map of the first quadrant, filtered by activities

**Table 3.** Activities performed in the quadrant after applying filters and its frequency.

| *Activity* | *Frequency* | *Relative frequency* |
|---|---:|---:|
| Assignment | 50646 | 29,48% |
| Closed | 29426 | 17,13% |
| Open | 27457 | 15,98% |
| Reassignment | 27077 | 15,76% |
| Caused By CI | 20769 | 12,09% |
| Communication with customer | 3607 | 2,10% |
| Pending vendor | 3445 | 2,01% |
| External Vendor Assignment | 2825 | 1,64% |
| Mail to Customer | 2556 | 1,49% |
| Reopen | 1046 | 0,61% |
| Resolved | 1015 | 0,59% |
| Communication with vendor | 967 | 0,56% |
| Vendor Reference | 555 | 0,32% |
| Analysis/Research | 427 | 0,25% |

**Quadrant II (15.2% of the cases)**

This quadrant contains the cases where the user contacts the service desk for a request for information, and there are not incidents. Figure 15 shows the relation between change cases and this type of incidents.

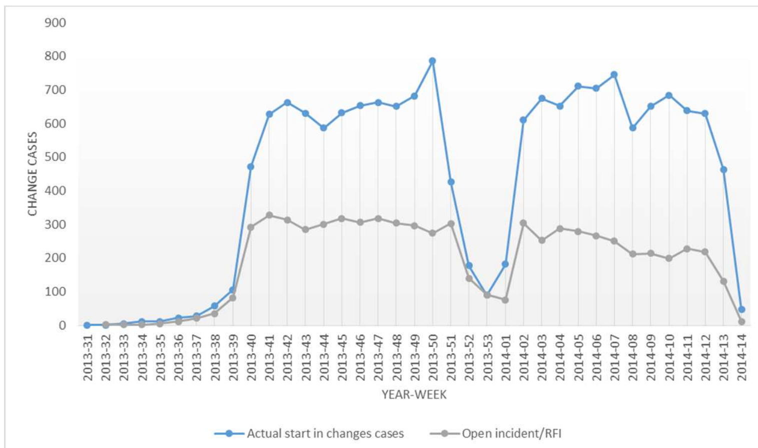



**Fig. 15.** Workload comparison between Service Desk and Changes Teams, expressed in cases opened every week, filtered by category as {Request for Information} and closure code as {Data, Hardware, Software, Other, Unknown, Referred, Operator error}

We modeled the process using the Flexible Heuristic Algorithm (Figure 16), that shows several parallel path product of the big amount of variants observed in the event log. In the data analysis we found there are only 33 activities (from the 39 of the original log), and 13 of them are performed in less than a 0.15% of the cases. We rebuild the model considering only the most common activities, and obtained the model shown in Fig. .

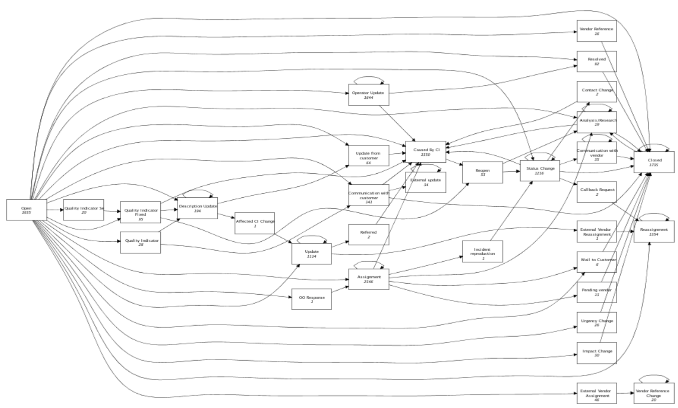


**Fig. 16.** Process map obtained using Flexible Heuristic Algorithm for the second quadrant data.

Using the same criteria as in quadrant I, the less common activities were filtered, as well the update activities. The resultant model contained just 12 activities that are shown in Table 4 and the resultant model in Figure. 17.

**Table 4.** Activities performed in the second quadrant after applying filters and its frequency.

| *Activity* | *Frequency* | *Relative frequency* |
|---|---|---|
| Assignment | 11430 | 27,23% |
| Reassignment | 9862 | 23,50% |
| Closed | 7483 | 17,83% |
| Open | 6790 | 16,18% |
| Caused By CI | 4642 | 11,06% |
| Reopen | 541 | 1,29% |
| Communication with customer | 429 | 1,02% |
| Resolved | 204 | 0,49% |
| External Vendor Assignment | 177 | 0,42% |
| Notify By Change | 163 | 0,39% |
| Quality Indicator Fixed | 136 | 0,32% |
| Analysis/Research | 117 | 0,28% |

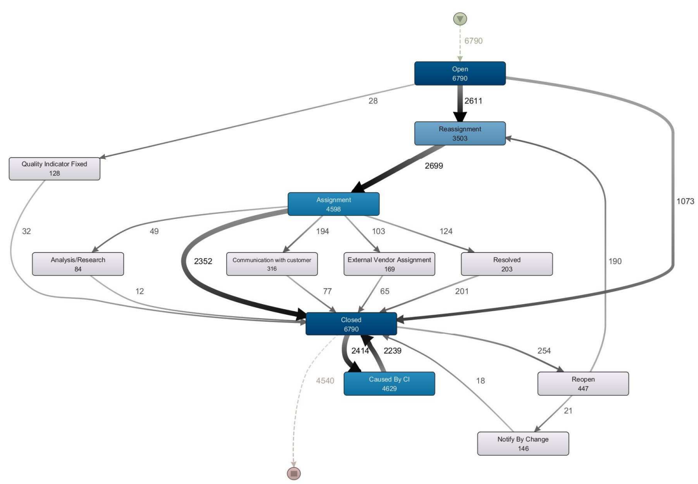



**Fig. 17.** Process map obtained using Disco software, for the second quadrant data, considering only the most frequent activities.

**Quadrant III (13.7% of the cases)**

This quadrant contains the cases where the user contacts the service desk to report a disruption, but is not an incident, is a user problem. Figure 18 shows the relation between change cases and this type of incidents.

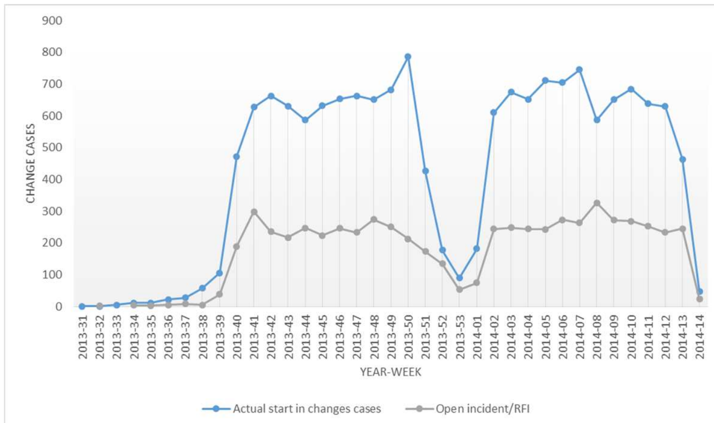



**Fig. 18.** Workload comparison between Service Desk and Changes Teams, expressed in cases opened every week, filtered by category as {Incident} and closure code as {No error - works as designed, Questions, User error, User manual not used}.

We modeled the process using the Flexible Heuristic Algorithm (Figure 19) that shows several parallel path product of the big amount of variants observed in the event log, this is the same observation of the second quadrant. In the data analysis we found there are only 37 activities, and 14 of them are performed in less than a 0.15% of the cases.

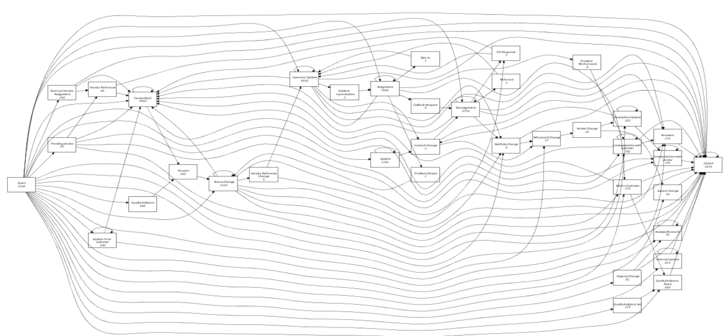


**Fig. 19.** Process map obtained using Flexible Heuristic Algorithm for the third quadrant data.

We rebuild the model considering only the activities with more than 0.15% of absolute frequency, and also filtering the update activities with the same criteria of the first two quadrants. The process map obtained is shown in Fig. 20 and the activities performed in Table 5.

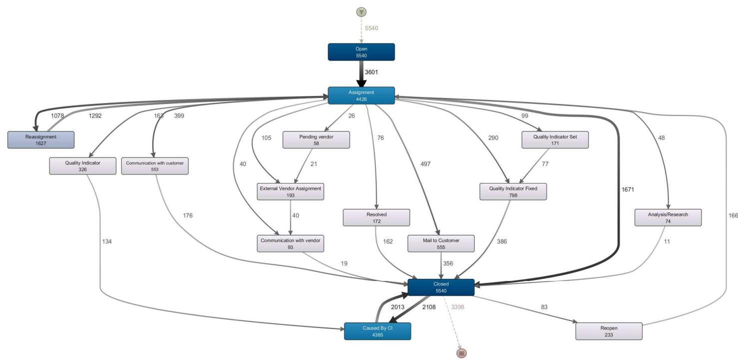



**Fig. 20.** Process map obtained using Disco software, for the third quadrant data, considering only the most frequent activities.

**Table 5.** Activities performed in the third quadrant after applying filters and its frequency.

| *Activity* | *Frequency* | *Relative frequency* |
|---|---:|---:|
| Assignment | 7936 | 25,41% |
| Closed | 5919 | 18,95% |
| Open | 5540 | 17,74% |
| Caused By CI | 4403 | 14,10% |
| Reassignment | 3759 | 12,03% |
| Quality Indicator Fixed | 880 | 2,82% |
| Communication with customer | 782 | 2,50% |
| Mail to Customer | 573 | 1,83% |
| Quality Indicator | 340 | 1,09% |
| Reopen | 260 | 0,83% |
| External Vendor Assignment | 200 | 0,64% |
| Quality Indicator Set | 175 | 0,56% |
| Resolved | 173 | 0,55% |
| Communication with vendor | 141 | 0,45% |
| Analysis/Research | 91 | 0,29% |
| Pending vendor | 63 | 0,20% |

**Quadrant IV (3.7% of the cases)**

This quadrant contains the cases where the user contacts the service desk for a request for information, and the case is indeed a question or a user problem. Figura 21 shows the relation between change cases and this type of incidents.

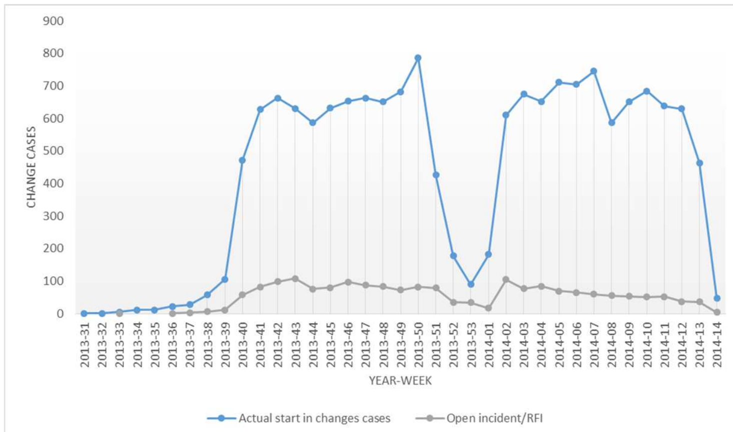



**Fig. 21.** Workload comparison between Service Desk and Changes Teams, expressed in cases opened every week, filtered by category as {RFI} and closure code as {No error - works as designed, Questions, User error, User manual not used}.

We modeled the process using the Flexible Heuristic Algorithm (Figure 22), and the result was the same that in the previous models, indicating several parallel path product of the big amount of variants observed in the event log.

This is the smallest quadrant, with just a 3.7% of the cases. We filter the data segment using the same criteria as in the previous quadrants, not considering the information update activities. The remaining activities were 23, where 7 of them had been performed in less than a 0.15% of the cases. With this second filter, the event log

The filtered model had 16 activities which frequency is shown in Table 6. The process map obtained is shown in Figure. 23.

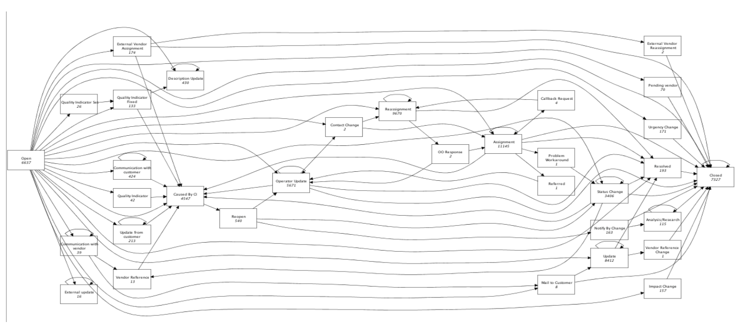


**Fig. 22.** Process map obtained using Flexible Heuristic Algorithm for the fourth quadrant data.

**Table 6.** Activities performed in the fourth quadrant after applying filters and its frequency.

| *Activity* | *Frequency* | *Relative frequency* |
|---|---|---|
| Assignment | 7936 | 25,41% |
| Closed | 5919 | 18,95% |

| | | |
|---|---|---|
| Open | 5540 | 17,74% |
| Caused By CI | 4403 | 14,10% |
| Reassignment | 3759 | 12,03% |
| Quality Indicator Fixed | 880 | 2,82% |
| Communication with customer | 782 | 2,50% |
| Mail to Customer | 573 | 1,83% |
| Quality Indicator | 340 | 1,09% |
| Reopen | 260 | 0,83% |
| External Vendor Assignment | 200 | 0,64% |
| Quality Indicator Set | 175 | 0,56% |
| Resolved | 173 | 0,55% |
| Communication with vendor | 141 | 0,45% |
| Analysis/Research | 91 | 0,29% |
| Pending vendor | 63 | 0,20% |

Using this characterization of the incident cases, and adding the demand prediction based in the change cases and previous information, the decision makers have information to perform a crew planning with a very low error.

Similar analysis can be done with the detailed information of CI impacted by changes and the incident cases reported, in order to have a very accurate prediction of the activities and complexity of them one week in advance.

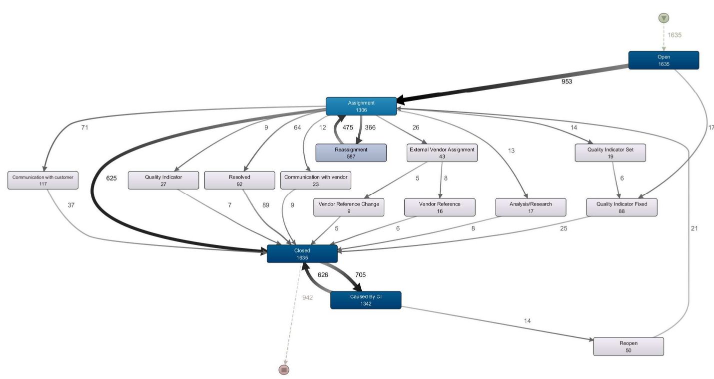



**Fig. 23.** Process map obtained using Disco software, for the fourth quadrant data, considering only the most frequent activities.

# 4 Change process review

This analysis is related with the change process and the average steps to resolution of a group of incidents. The idea of this analysis is to provide a review about the behavior of the service levels after each change execution process that has been implemented. For the description of the change process in Rabobank, to start, an initial analysis of the log was done, identifying the attributes and its values.

The main attributes identified are: Activity, Change Type, # Related incidents, # Related Interactions, CAB- approval needed, CI Name, CI Subtype, CI Type, Emergency Change, Originated From, Risk Assessment and the WBS affected.

The focus of the main analysis will be centered on three main attributes:

1. For the Activity attribute, 9 different values are identified: Change record Open Time, Change record Close Time, Requested End Date, Planned Start, Planned End, Actual Start, Actual End, Schedule Downtime Start and Schedule Downtime End.
2. For the Change type attribute, 6 main types are identified: Change Management, Master Change, Master Change Roadmap, Release Type, Standard Activity Type and Standard Change type.
3. For the CI Type attribute, 13 values are identified: Phone, Application, Application Component, Computer, Database, Display Device, Hardware, Network Components, no type, Office Electronics, Software, Storage and Sub Application.

Based on the Change Type Attribute, an analysis was made to discover which of these six types are more relevant in the event log, discovering that only three types are relevant and that they represent 91% of the log, so from now on the analysis will be done with only the following types: Release Type, Standard Activity Type and Standard Change type.

*Process models*

For each change type, we discover the process model using DISCO. Bellow, we show the Figures 24, 25 and 26 of each process model.

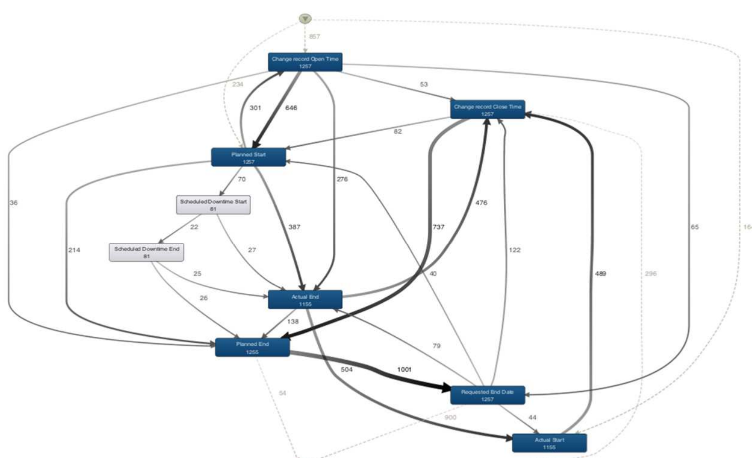



**Fig. 24.** Releases process model.

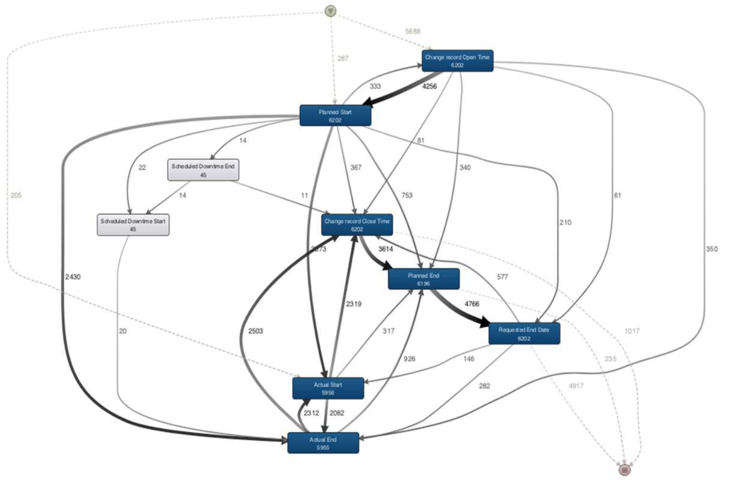


**Fig. 15.** Standard activity type process model.

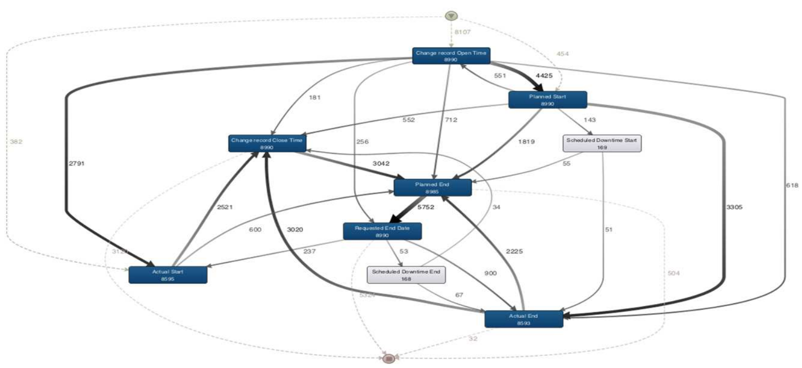


**Fig. 26.** Standard change type process model.

Based on these previous models, we identified some similarities and differences between them, which are shown in the following table 7.

**Table 7.** Change type general information.

| Change type | Similarities | Differences |
|---|---|---|
| Releases | 9 activities<br>*Scheduled downtime start* and *Scheduled downtime end* present low frequency. | 3 last in cases activities: *Requested end date*, *Change record close time*, *Planned end.* |
| Standard activity type | 9 activities<br>*Scheduled downtime start* and *Scheduled* | 3 last in case activities:<br>*Requested end date*, *Change record close* |

| | | |
|---|---|---|
| | *downtime end* present low frequency. | *time*, *Planned end*. Few cases after finish after *actual end* activity are executed. No cases finish with *actual end* activity. |
| Standard change type | 9 activities<br>*Scheduled downtime start* and *Scheduled downtime end* present low frequency. | 4 last in case activities:<br>*Requested end date*, *Change record close time*, *Planned end*, *Actual end*. Few cases after finish after *actual end* activity are executed. |

*Comparing Actual Start & End time vs. Planned Start & End time*

Based on the values of the activities, four main activities are identified as the source of the estimated and real times.

*Real Time*

For the real time, two activities were filtered, the actual start and the actual end. Executing this filter 87% of the cases were included, and are divided in four variables:

- 10027 (Start-End have valid values)
- 5675 (Start and End have value 0)
- 4 (Only have start)
- 1 (Only have end)

Of these four variables, the only valued for analysis is the first one, in this case only 10027 cases are described. For these values the median is 56.8 minutes and the mean is 32.8 hours.

*Planned Time*

For the planned time, two activities are filtered, the planned start and the planned end. Executing this filter 83% of the cases were included, and are divided in three variables:

- 15072 (Start-End have valid values)
- 1382 (Start and End have value 0)
- 13 (Only have start)

Of these three variables, the only valued for analysis is the first one, in this case only 15072 cases are described. For these values the median is 22.5 hours and the mean is 4.3 days. After this description, each extracted log was split into each of the three main change types so a comparison could be made between the planned time and the real time. Following are the three main change types and the graphics obtained.

*Change type: Releases*

The Figure 27 shows the graphic associated with Releases. In this change type, we note there are only few cases where a marked difference, between the planned time and the real time to deploy each release, is presented.

The releases where the difference appears are: *Release 1*, *Release 3*, *Release 6*, *Release 9* and *Release 10*, shown in Table 8. Exploring in more detail these five releases, we discovered the following characteristics shown in table CC. A more detailed analysis is shown at the end of the section.

The main conclusion in this analysis, is that the releases, which include changes towards the end of 2013 and beginning of 2014, have better real time vs. estimated time, than the previous releases. This reflects that the teams are not only estimating better times, but they are executing the releases in the estimated times. The way the teams are working should continue this way because the improvement is significantly visible towards 2014.

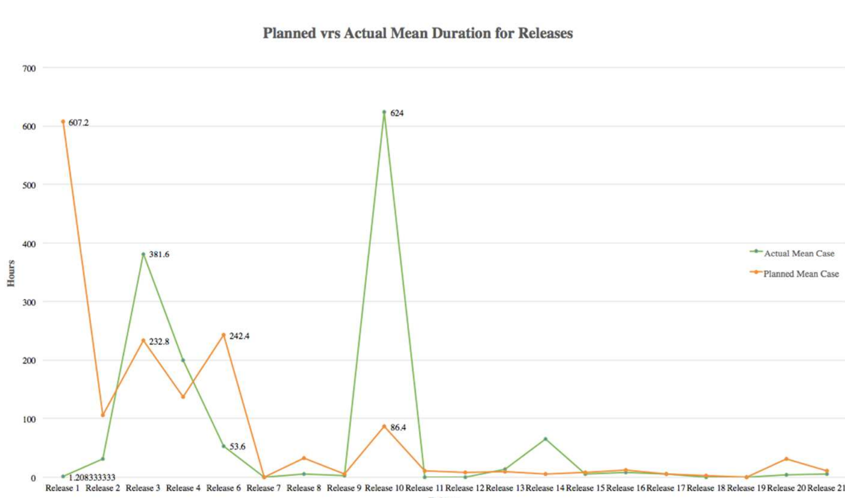



**Fig. 27.** Releases change type graphic

**Table 8.** Releases characteristics

| | 1 | 3 | 6 | 10 |
|---|---|---|---|---|
| **Characteristic** | | | | |
| Event | 201 | 950 | 1607 | 12 |
| Cases | 29 | 136 | 232 | 2 |
| Activity | 7 | 7 | 7 | 7 |
| Mean case duration | 30.7 d | 20.6 d | 13 d | 23.3 d |
| Start | 09/06/13 | 9/9/13 | 4/9/13 | 2/10/13 |
| End | 28/3/14 | 28/3/14 | 31/3/14 | 28/10/13 |
| CI Type | Application 86% | Application 100% | Application 69% | Application 100% |
| Emergency | No | No | No | No |

| | | | | |
|---|---|---|---|---|
| case | | | | |
| Originated from | Problem 100% | Problem 100% | Problem 98% | Problem 100% |
| Risk assessment | Minor Change 100% | Minor Change 100% | Minor Change 100% | Minor Change 100% |

*Change type: Standard activity type*

The Figure 28 shows the graphic associated with *Standard activity type*. Here, we note more cases where a marked difference is presented, between the planned time and the real time, to deploy each change in this category.

The Standard activity types where the differences appear are: *9*, *15*, *17*, *33*, *36*, *38* and *49*, shown in Table 9. Exploring in more detail these seven change type, we discovered the following characteristics shown in table CC. A more detailed analysis is shown at the end of the section.

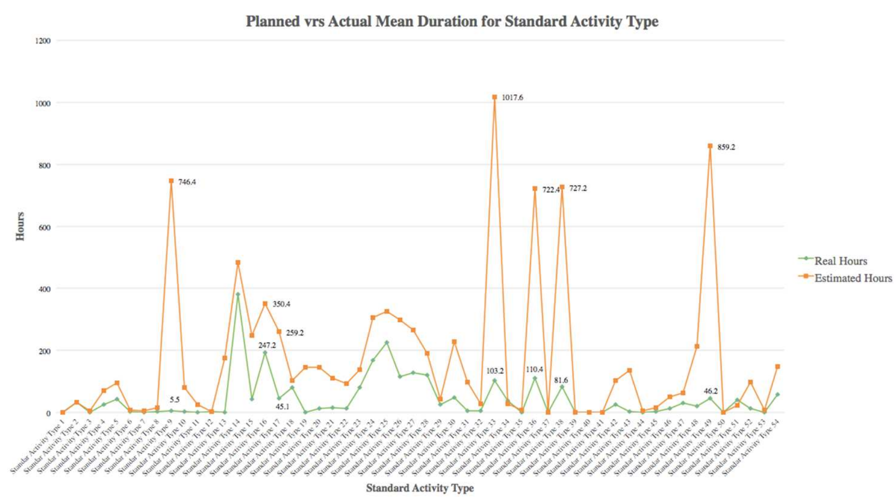



**Fig. 28.** Standard activity type graphic

**Table 9.** Standard activity characteristics

| | 9 | 15 | 17 | 33 | 36 | 38 | 49 |
|---|---|---|---|---|---|---|---|
| **Characteristic** | | | | | | | |
| Event | 9560 | 105 | 182 | 674 | 87 | 42 | 229 |
| Cases | 1370 | 15 | 26 | 102 | 13 | 6 | 33 |
| Activity | 9 | 7 | 7 | 7 | 7 | 7 | 7 |
| Mean case duration | 5.2 d | 35.1 d | 26.7d | 59.4 d | 41.2 d | 34.2 d | 43.7 d |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Start | 24/09/13 | 18/9/13 | 2/12/13 | 26/9/13 | 10/9/13 | 6/2/13 | 19/9/13 |
| End | 31/3/14 | 31/3/14 | 31/3/14 | 27/2/14 | 17/2/14 | 13/3/14 | 31/3/14 |
| CI Type | Network Components 87%<br>Computer 10% | Computer 93%<br>Database 7% | Computer 100% | Phone 89%<br>Network Components 10%<br>Applications 1% | Network Components 100% | Network Components 67%<br>Phone 33% | Computer 100% |
| Emergency case | No | No | No | No | No | No | No |
| Originated from | Incident 100% | Incident 100% | Problem 100% | Problem 100% | Incident 100% | Incident 100% | Incident 100% |
| Risk assessment | Minor Change 99%<br>Mayor Business change 1% | Minor Change 100% | Minor Change 100% | Minor Change 100% | Minor Change 100% | Minor Change 100% | Minor Change 100% |

The main conclusion from the previous analysis is that the standard activity types do not reflect a significant decrease like the releases in time, and several high differences between planned and real time are still shown towards the highest standard activity types. It is important to notice that in this type of changes, it is not simple to say that the higher the standard activity type is, the most recent in time, because no time order is visible. In general, the estimated times are considered good, but still work should be done to make better estimation and real times.

*Change type: Standard change type*

The Figure 29 shows the graphic associated with *Standard change type*. This is the change type that has the highest number of cases where there is variance between planned and real time.

The Standard change types where the differences appear are: *1, 16, 26, 57, 61, 125, 130, 147 and 156*, shown in Table 10. Exploring in more detail these nice cases, we discovered the following characteristics shown in table CC. A more detailed analysis is shown at the end of the section.

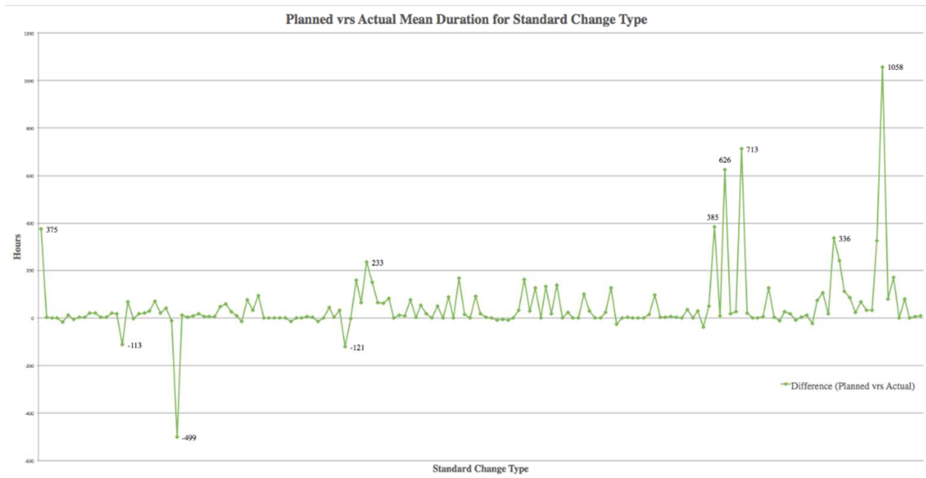



**Fig. 29.** Standard change type graphic

**Table 10.** Standard change characteristics

| | 1 | 16 | 26 | 57 | 61 | 125 | 127 | 130 | 147 | 156 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Characteristic** | | | | | | | | | | |
| Event | 26 | 621 | 14 | 14 | 44 | 7 | 10 | 7 | 348 | 371 |
| Cases | 4 | 89 | 2 | 2 | 6 | 1 | 2 | 1 | 54 | 53 |
| Activity | 7 | 7 | 7 | 7 | 9 | 7 | 5 | 7 | 7 | 7 |
| Mean case duration | 15.1 d | 30.8 d | 21.7 d | 21.8 w | 19.1 d | 50.1 d | 30 d | 30.8 d | 37.4 d | 85,5 d |
| Start | 6/1/13 | 8/10/13 | 20/9/13 | 24/9/13 | 21/1/13 | 12/12/13 | 17/1/14 | 24/1/14 | 6/9/13 | 25/9/13 |
| End | 12/2/14 | 16/1/14 | 28/2/14 | 8/3/14 | 27/2/14 | 31/1/14 | 20/2/14 | 24/2/14 | 30/3/14 | 20/3/14 |
| CI Type | Computer 46%<br>New. Components 27%<br>no type 27% | Computer 100% | Database 100% | Hardware 100% | Computer 69%<br>Hardware 32% | Computer 100% | Application 100% | Application 100% | Application 91%<br>Computer 5%<br>Sub-application 4% | Computer 79%<br>Application 21% |
| Emergency case | No | No | No | No | No | No | No | No | No | No |
| Originated from | Problem 73%<br>Incident 27% | Incident 100% | Problem 100% | Problem 100% | Problem 80%<br>Incident 20% | Problem 100% | Problem 100% | Problem 100% | Problem 100% | Problem 100% |
| Risk assessment | Minor Change 100% | Minor Change 100% | Minor Change 100% | Minor Change 100% | Minor Change 100% | Minor Change 100% | Business change 50%<br>Minor Change 50% | Minor Change 100% | Minor Change 100% | Minor Change 100% |

The main conclusion from the previous analysis is that the standard change type, the same as the standard activity types do not reflect a significant decrease like the releases in time, and several high differences between planned and real time are still shown. It is also important to notice that in this type of changes, it is not simple to say that the higher the number of standard change types, the most recent in time, because no time order is visible. It is important to see that in the standard change types most of them have positive differences indicating good real vs. planned times, meaning that the ending of the real work is done within the estimated times, but work should be still done to lower both the real and estimation times in general.

*Characterization of exceptional cases*

Comparing the Tables 4, 5 and 6 and the information related with the characteristics of each change type, we highlight the following points:

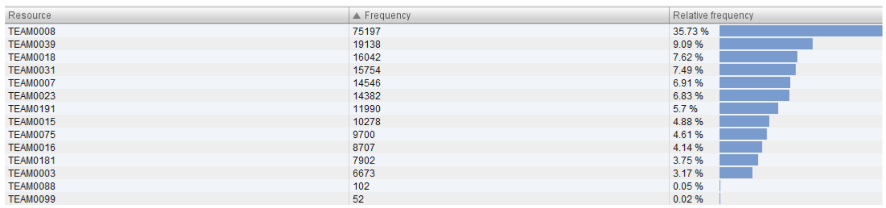
- The amount of activities per change type is similar between in each change type.
- It can be seen that the mean case duration for the cases in the Release Change Type is lower, compared with the Standard Activity Type and the Standard Change Type. The Standard Change Type has the highest mean case duration (85, 5 d).
- The CI Types most affected are: computer, network components, application and sub-application. In Standard Activity Type, two more CI Types appear: Phone and Database. For the Standard Change Types, Database also appears, and Hardware is another one listed. For the Releases, just the CI Type Applications appear.
- Considering the characteristic Originated from, in the Release Change Type, all the changes were caused from a problem. In the Standard Activity Type, most changes were produced from an incident. In the Standard Change Type, problem is the main originator; however, incident as a cause appears in three of them.
- In the case of Risk Assessment, the Minor Change is the most common type in the three exceptional cases. Just in one case of the Standard Change Types, a Business Change appears with a 50%.

## 5 Additional analysis

In addition to previously answered questions, is interesting to analyze the relation between the teams that participate in the interaction, incident and change management.

*Team analysis*

For this analysis, we consider the complete cases (have the open activity eventually followed by the closed activity), between October first 2013 and April first 2014. We focus on the four main CI Types: Applications, Sub-applications, Computer and Storage. With this information, we discover the frequency of the resources involved in the process. As a first approach, we see that the Team0008 is the team that has the higher workload compared with the other teams. The following Figure 30 shows the main teams involved.



| Resource | Frequency | Relative frequency |
| --- | --- | --- |
| TEAM0008 | 75197 | 35.73 % |
| TEAM0039 | 19138 | 9.09 % |
| TEAM0018 | 16042 | 7.62 % |
| TEAM0031 | 15754 | 7.49 % |
| TEAM0007 | 14546 | 6.91 % |
| TEAM0023 | 14382 | 6.83 % |
| TEAM0191 | 11990 | 5.7 % |
| TEAM0015 | 10278 | 4.88 % |
| TEAM0075 | 9700 | 4.61 % |
| TEAM0016 | 8707 | 4.14 % |
| TEAM0181 | 7902 | 3.75 % |
| TEAM0003 | 6673 | 3.17 % |
| TEAM0088 | 102 | 0.05 % |
| TEAM0099 | 52 | 0.02 % |

**Fig. 30.** Team analysis filter

Applying this filter, we obtain 81% of the cases, which correspond to the 14 most frequent teams. Considering this team data, we decided to obtain the information related with the relative frequency and the mean case duration. The Table 11 shows this information about each team.

**Table 11.** Team information

| Team | Relative frequency | Mean case duration |
| --- | --- | --- |
| TEAM0003 | 3,17% | 7,8 d |
| TEAM0007 | 6,91% | 4,9 d |
| TEAM0008 | 35,73% | 23,8 h |
| TEAM0015 | 4,88% | 6,4 d |
| TEAM0016 | 4,14% | 4 d |
| TEAM0018 | 7,62% | 5,1 d |
| TEAM0023 | 6,83% | 11,5 h |
| TEAM0031 | 7,49% | 10,5 h |
| TEAM0039 | 9,09% | 9,8 h |
| TEAM0075 | 4,61% | 35,6 h |
| TEAM0088 | 0,05% | 4 d |
| TEAM0099 | 0,02% | 7,9 h |
| TEAM0181 | 3,75% | 32,3 h |
| TEAM0191 | 5,7% | 13,4 h |

With this analysis, we decided to apply organizational metrics from the process mining tool Prom 6.3, to find the relationship between these main teams.

*1. Handover of work*

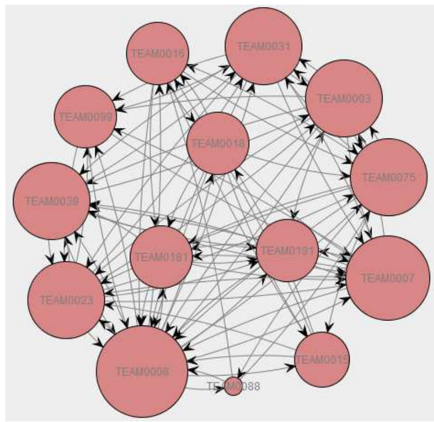



**Fig. 31.** Team analysis: Handover of work

As we can see in Figure 31, with this metric we discovered that most of the teams work directly with each other, but mainly with the TEAM 0008.

*2. Working together*

With this metric, we want to discover if there are teams that attended particular cases. In the Figure 32, we can see that there is no separation between the teams according to the CI Types.

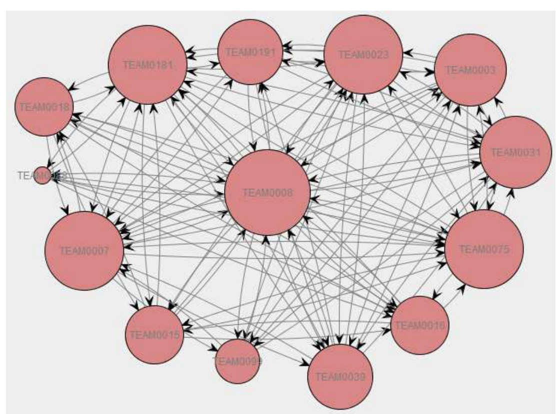



**Fig. 32.** Team analysis: Working together

*3. Doing similar tasks*

Because no work teams were discovered, the similar task metric can help us identify the existence of roles in the process. Figure 33 also shows that it is not possible to determine roles.

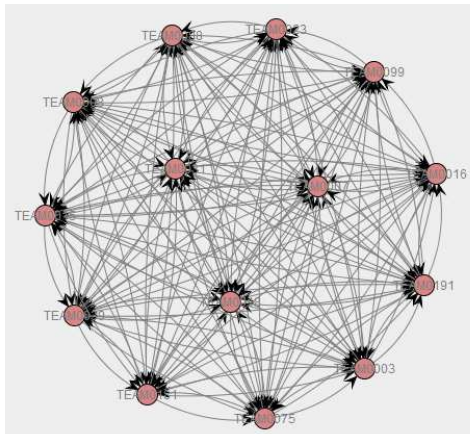


**Fig. 33.** Team analysis: Doing similar tasks

*4. Subcontracting*

With the subcontracting metric is possible to see that all the main teams involved in the process have a close relation with the TEAM0008, which is the only team that has subcontracting tasks with all the rest of the teams. TEAM0088 does no any type of subcontracting. Figure 34 shows the resulting diagram. We decided to characterize the following teams to see if any patterns of participations were discovered.

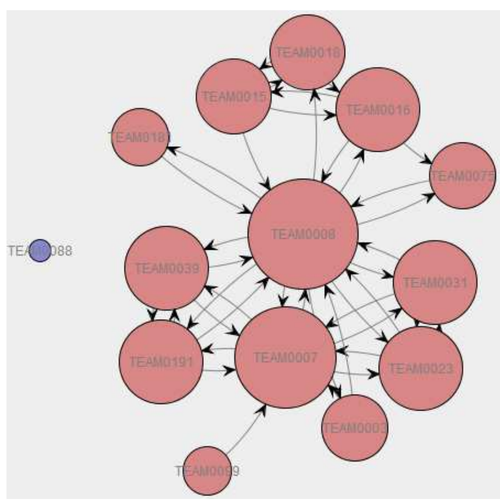



**Fig. 34.** Team analysis: Subcontracting

1. *Group 1 TEAM0008*
   The TEAM0008 is the team that does the open activity most (around 30%), having participation in case with the other groups.

2. *Group 2 TEAM0007*
   The TEAM0007 does not open cases but closes a lot of cases. Its participation is more towards the ending part of the process.

3. *Group 3 TEAM0015, TEAM0016, and TEAM0018*
   TEAM0015 does not open cases, but the behavior of TEAM0016 and TEAM0018 is very similar to TEAM0008. This three teams focus more on the assignment and reassignment activities that opening or closing activities.

4. *Group 4 TEAM0039 and TEAM0191*
   These two teams do not open cases; they only close them and also work more towards the assignment and the status change of the cases.

5. *Group 5 TEAM0031 and TEAM0023*
   These two teams have similar behavior as the ones in group 4.

6. *Relationships of groups 3, 4 and 5 with group 1 and group 2*
   The relationship between the group 1 and group 3 is less evident, having both a clearly active participation in the opening and closing activities of the cases. Three teams open cases actively (TEAM0008, TEAM0016 and TEAM0018), and one team closes cases frequently (TEAM0015). It is clear to say that TEAM0015 is the team that has less similarity with TEAM0008 from group 1. The relationships between the group 1 and groups 4 and 5 y basically that TEAM0008 opens most of the incident cases for teams 0023, 0031, 0039, and 0191. Group 2 has a team that participates more in the middle and closing activities of the process, so the relationship with group 4 and 5 is higher that its relationship with group 3.

From the above review we can see that some teams have a clearly more efficient way to work towards the first activities in the incident process, meanwhile other teams have a more active participation towards the middle and closing activities of the process.

## 6 Conclusions

Based on the analysis of the available information, we created a prediction model for service desk workload, we explored the incidents and request for information categories, considering the number of opened changes cases and the number of received calls by the service desk, during the previous two weeks of the prediction.

The results obtained with this model show that the correlation detected was high and the relative absolute error percent low. Also, we develop a prediction model that considers the number of incidents related with the closed change cases. The results of this model show a high precision in the prediction of the number of incidents based on the incidents and closed change cases from the previous week and how this cases impacted each CI Type.

To achieve a more accurately workload estimate, the calls received by the service desk were categorized into four quadrants, identifying activities, their frequency, and the process map followed by them in each case. Each quadrant has a different impact according to the predictive models. With this, it is possible to do a better planning regarding the available resources and their capabilities.

For the change management process, we provided an analysis about the behavior of the service levels after each change execution process that was implemented. We identified the attributes related, and focused in three main attributes: activity, change type, and CI Type. Based on the Change Type Attribute, an analysis was made to discover which are more relevant in the event log, determining that only three types are relevant and that they represent 91% of the log. Considering this, we realized a further analysis considering only these types: Release Type, Standard Activity Type and Standard Change type. The log was split into each of the three main change types and a comparison was conducted considering the variance between the planned time and the real time. We found that the Releases Type Changes, present an improvement, towards 2014, in real time versus estimated time, than the previous releases. This reflects that the teams are not only estimating better times, but they are executing the releases in the estimated times. In the case of the Standard Activity Type and the Standard Change Type, these types do not reflect a significant decrease like the releases in time, and several high differences between planned and real times are found.

In addition, we analyze the relation between the teams that participate in the interaction, incident and change management. Using organizational mining metrics, we found that no work teams were discover, and it is not possible to determine roles according to the four main CI Types selected. A group characterization was proposed, and we conclude that some teams have a clearly more efficient way to work towards the first activities in the incident process; meanwhile, other teams have a more active participation towards the middle and closing activities of the process.

## References

1. Van der Aalst, W. M.: Discovery, Conformance and Enhancement of Business Processes. Springer, Heidelberg (2011)
2. van Dongen, B. F., de Medeiros, A. K. A., Verbeek, H. M. W., Weijters, A. J. M. M., & Van Der Aalst, W. M.: The ProM framework: A new era in process mining tool support. In Applications and Theory of Petri Nets 2005 (pp. 444-454). Springer Berlin Heidelberg (2005)
3. Günther, C. W., & Rozinat, A.: Disco: Discover Your Processes. In BPM (Demos) (pp. 40-44), (2012)
4. Han, J., & Kamber, M.: Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan kaufmann (2006)
5. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H.: The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18 (2009)