Service Desk and Incident Impact Patterns Following ITIL Change Implementation

Pierre Buhler, Robert O' Callaghan, Soline Aubry, Danielle DeJoy, Emily Kuo, Natalie Shoup, Inayat Khosla, Mark Ginsburg, Nicholas Hartman, and Nicholas S. McBride

CKM Advisors, 711 3^{rd} Ave. Suite 1806, New York, NY, 10017 USA
{pbuhler,rocallaghan,saubry,ddejoy,ekuo,nshoup,ikhosla,mginsburg, nhartman,nmcbride}@ckmadvisors.com

Abstract. As part of the 2014 Business Process Intelligence Challenge we analyzed ITIL event logs surrounding the Change Management process to build a model to understand and predict the impact of Change implementation on Incident management. 13 impact patterns were identified and their key metrics calculated. Multinomial regression modeling indicated several segmentation criteria of Changes that significantly influenced impact pattern type and whether the impact was favorable or unfavorable to downstream Service Desk and IT Operations. Performance analysis of associated IT service products revealed that 7% of product lines are simplifying their Incident resolution process through Change implementation and 2% are introducing additional complexity. 91% show no significant change in complexity.

Key words: Mining of Business Processes from Event Logs; Performance Measurement of Business Processes; Process Discovery; Resource Allocation in Business Processes

1 Introduction

IT process management surrounding the Information Technology Infrastructure Library (ITIL) framework is an area ripe for optimization through the use of digital analytics [1], [2], [3]. We leveraged these techniques on data provided by Rabobank as part of the 2014 Business Process Intelligence Challenge [4], with the objective of understanding, predicting and preventing additional workload on the Service Desk (SD) and IT Operations (ITO) teams as a result of Change implementation. We also developed metrics that measure improvement in service levels through Change implementation by product managers. Integration of data across three sub-streams—the Change, Interaction and Incident processes—was required to perform the analyses.

The lack of defined connectivity between Change records and resulting Interactions and Incidents necessitated the isolation of a sample group in order to associate Interactions and Incidents with specific Changes, so as to better identify impact patterns. These patterns were analyzed and 13 mutually exclusive collectively exhaustive impact pattern types were identified to characterize the impact of a change on a service component. These change patterns would be used as a means of grouping known changes and classifying future impact predictions.

Once these patterns were established, each change was analyzed and labeled with the appropriate change pattern. To predict the change patterns of future changes we proceeded to evaluate possible predictive drivers of impact based on known Change information. Significant standalone drivers of impact were identified. Two separate predictive models were created, one based on the distribution of change patterns across the known data and the second leveraging the known information on predictive drivers. The data set was split into a training set and a test set on which the two predictive models were trained and tested. The performances of these predictive models were then tweaked, analyzed and assessed for accuracy against each other and against baseline values. Both predictive models performed substantially better than the defined baseline. This demonstrated the potential for optimization of process management through leveraging predictive digital analytics.

Lastly, we evaluated product manager performance to determine if Change implementation led to improved service levels, both incrementally and over time. Additional observations on Incident management process improvements are also discussed.

1.1 Initial Understanding of the BPIC Dataset

Four data tables were provided. Three of these files—Interaction Records, Incident Records and Change Records—contain case metadata (Case ID, related Configuration Items and Service Components, Open, Closed and Handle Times, related activity across processes). The fourth file—Incident Activity Records—contains Incident event logs, with each row capturing a particular step toward Incident closure.

The dataset encompassed data extracted within the six-month period of October 2013–March 2014. Because the data were extracted based on closure time, an edge effect was observed. Cases that had been opened before 31 March, but not yet resolved, were excluded from the dataset. In order to prevent this from skewing our results, we excluded the final 17 days of the dataset from our timeframe[1]. As such, the window for our analysis is 1 October 2013–14 March 2014.

The original .CSV files were loaded into an R environment and converted to appropriate data types, such as standardized timestamp formats, for analysis. Data were then loaded on to a purpose-built MySQL database. Analyses were performed using RStudio Server and R version 3.0.2 interfacing with the MySQL database through the DBI and RMySQL packages. Additionally, MySQL Workbench was used to query and build tables within the database. Processes were visualized and summary calculations rendered using Fluxicon Disco.

[1] 95% of incidents are closed within 17 days. By excluding the last 17 days of the dataset we were able ensure that the bulk of Incidents opened within the timeframe are represented in our analysis.

After onboarding the data, we traced the process flow of the dataset. Fig. 1 illustrates the process as detailed in the documentation [5].

We used the case identifier fields (Interaction ID, Incident ID and Change ID) to link elements between the different sub-streams. Examination of these identifiers revealed inconsistent coverage of linkages across the four data tables. By adding linkage data from other tables we were able to increase the number of Incidents with unique links to Interactions from 92.4% to 99.7%.

Based on this process, we found 1,568 Interactions and 868 Incidents that were explicitly linked to their causal Changes. This represents less than 1% of Incidents and Interactions within the dataset. Based on our experience with similar datasets from prior work, it is common for such explicit linkages to be absent from ITIL process data unless there is an enforced effort to track these relationships. We also learned that explicit linkages were only introduced when this link is 'obvious to the technician' [6]. This left the probability that many Incidents and Interactions caused by Change activity remained unidentified. In conjunction with explicitly linked Incidents, these unidentified impact events are vital to understanding the comprehensive impact of Change implementation on the SD and ITO.

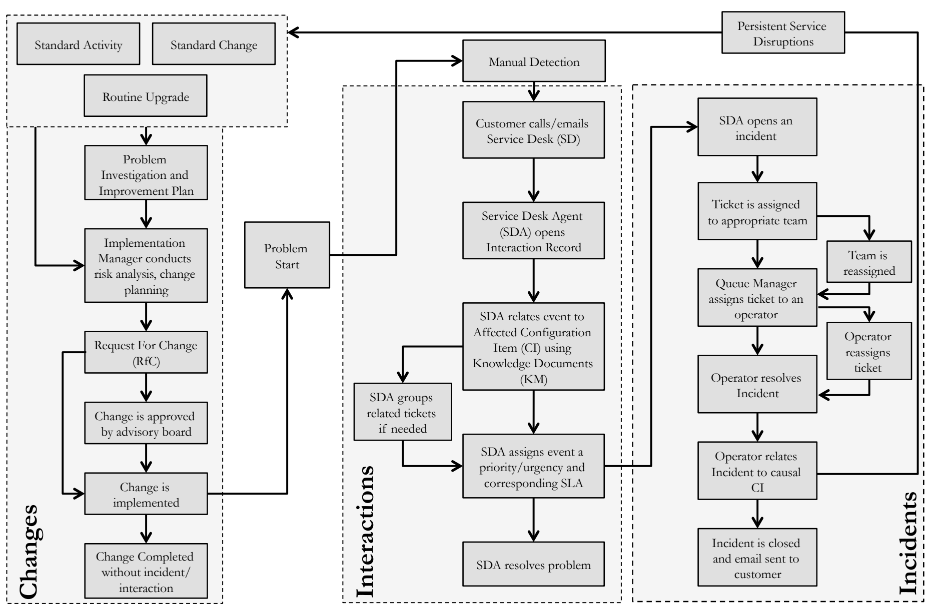

Fig. 1. Illustrative process map of Change, Interaction and Incident sub-streams.

We developed a more comprehensive system of linking to include additional impact events not explicitly linked. The Service Component (SC) corresponds to a particular service offering within the bank, while the Configuration Item (CI) Name corresponds to a more granular asset identity. These two entities are related to Incidents either in an 'affected' relationship (the Incident affected the entity) or a 'caused by' relationship (the Incident was caused by the entity). An analysis of the 868 Incidents that were explicitly linked to Changes showed that the causal SC of the Incident matched the SC of the Change in 44% of cases, while the causal CI Name matched in only 18% of the cases. Similarly, the affected SC matched the SC of the Change in 33% of cases, while the affected CI Name matched in 15% of cases. We therefore concluded that SC would serve as the best means of linking Changes to resulting downstream activity for our exploratory sample. Focusing on SC allowed a natural alignment of our analysis with the product lines as they exist in the business and provided a means of linking otherwise unidentified impact events to Changes through SC matching. The Caused By SC and Affected SC columns were used as primary identifiers in the Incident and Interaction records respectively.

2 Impact Identification

2.1 Isolation of Change Events

Initial study of impact patterns showed that most ITIL activity, including Change implementation, occurred during working hours. 85% of Changes had a start time between 07:00 and 18:00 on weekdays. 99% of Interactions and Incidents were opened between the hours of 08:00 and 17:00 on non-holiday weekdays. Given this, we expect the impact patterns to be similarly aligned with working hours. Accordingly, when normalizing the timing of impact events against a Change, only the number of working hours between the Change and the impact event were counted.

17,172 Changes were started across 280 SCs within the time frame of our dataset. This includes many instances of multiple Changes occurring on the same SC in a short period of time. To enable the linkage of downstream events to individual Changes, we isolated Changes where no other Changes occurred in the same SC within a specified isolation time window. This was repeated for time windows ranging from ± 1–10 days from Change start[2]. In order to identify the proper isolation time window, we performed sensitivity testing to determine whether the number of Change-SC pairs was not overly sensitive to the time window used. This relationship was largely linear around the isolation window of three workdays. This combined with a visual inspection of typical impact pattern duration allowed us to conclude that a buffer of three days would be a reasonable timeframe for selecting the relevant set of observations on which to perform further analysis. Incidents and Interactions were then linked to a Change if they occurred within that Change's established time window.

[2] Change start time refers to the Actual Start Time given in the Change records table, or the Planned Start Time if Actual Start Time was not populated.

Table 1. The number of Changes, Interactions, and Incidents within the same Service Component that were extracted using isolation windows of 1–10 workdays. An isolation of three working days was chosen for further analysis.

	Isolation Window (workdays)									
	1	2	3	4	5	6	7	8	9	10
Changes	2,007	1,223	900	664	505	400	343	304	263	228
Interactions	6,561	5,014	4,853	3,846	3,329	2,448	2,039	1,927	1,808	1,714
Incidents	20,491	15,998	14,657	11,697	9,573	7,528	6,391	6,244	5,943	5,143

This set of Change-SC pairs isolated by three workdays is the set on which we sought to initially identify case studies of new impact events that had not been explicitly linked to Changes. It consists of 900 Changes representing 5.2% of all Changes within the timeframe, but accounting for 11.6% of the Incidents and 10.2% of the Interactions.

2.2 Impact Pattern Case Studies

In order to gain specific insight into the impact that Change implementation has at the SD and ITO for predictive modeling, we looked at the volume of Interactions and Incidents preceding and following individual Changes over a 27 working hour period (3 work days). Here we present examples of four impact patterns that were observed.

Acute Impact Change C00000589 is an example of a single Change impacting multiple SCs in divergent patterns. Fig. 2A shows an acute impact on SC WBS000152 with Interactions and Incidents rapidly spiking within 1 hr of Change start, then slowly decreasing. Conversely, Fig. 2B shows the impact pattern from the same Change on SC WBS000167, beginning more gradually after Change start (but before Change end[3] at t = 9 hrs), returning to baseline levels at t = 17 hrs. Future analysis could assess whether similar ITO teams handled the Incidents associated with these impacts across the two SCs and whether it is likely that they fall within the same area of management.

Extended Impact Fig. 2C depicts the long-lasting impact of Change C00015800 on the SC WBS000335 workload at both the SD and ITO. The volume of Interactions starts increasing 2 hrs after the end of the Change, followed by an increase in Incidents 3 hrs later at t = 6 hrs. The volume of new Interactions and Incidents remains elevated up to 27 working hours later.

Problem Resolution Change C00013569 depicted in Fig. 2D on SC WBS000006 originated from a 'problem'. This means that a certain number of Incidents were reported on the same issue, resulting in an investigation followed by Change implementation to correct the issue [5]. A constant volume of Interactions and

[3] Change end time refers to the Actual End Time given in the Change records table, or the Planned End Time if Actual End Time was not populated.

Incidents precedes the Change. At t = 5 hrs both new Interaction and Incident volume have dropped to zero, where they remain at a new, lowered steady state with only an occasional Interaction. This example illustrates a successful Change implementation leading to the resolution of a particular problem and effective decrease in both SD and ITO workload.

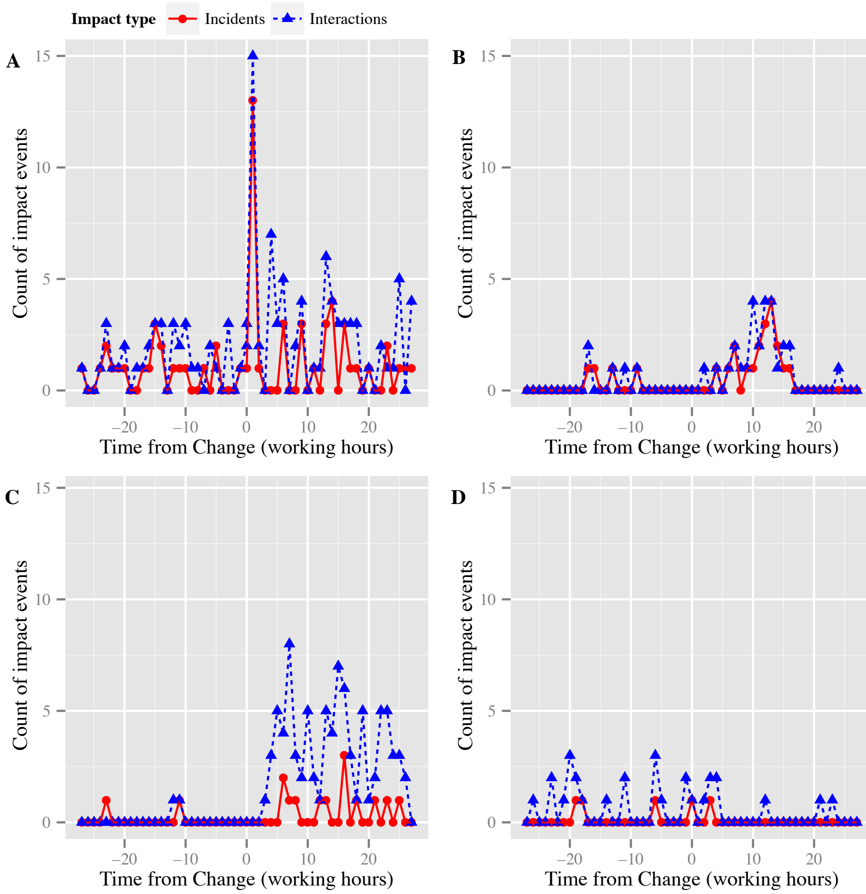

Fig. 2. Example Change impact pattern on the Service Desk (Interactions) and IT Operations (Incidents): **A.** Acute Impact; **B.** Delayed Impact; **C.** Extended Impact; **D.** Problem resolution.

In all cases shown here the elapsed time between Change implementation and the start of the impact pattern was correlated either to the 'actual start' or the 'actual end' of the Change. We chose to focus on Change start in order to be inclusive of any impact events. The knowledge that impact patterns can vary considerably between SCs—even those stemming from the same Change—will help direct further study. Impact patterns identified here will help inform future modeling endeavors by allowing us to fit parameters to their expected shapes and look for similar patterns surrounding Changes, with the added potential of identifying impacts across SCs.

3 Pattern Classification

In order to accurately predict the impact on the SD and ITO, we divided the data into separate datasets for Interactions and Incidents. Change-SC pairs in either dataset that exhibited low activity levels, i.e., less than 4 impact events opened either before or after Change start (t_0), were removed from the analysis. This left us with 8,766 and 5,833 Change-SC pairs in the Interaction and Incident datasets respectively.

3.1 Steady State Definition

The original BPIC statement refers to a 'steady state' to which the level of Interactions/Incidents is expected to return following perturbation by a Change impact [4]. Although there is no natural 'steady state' for a system exhibiting Poisson arrival of various perturbations, we sought a working definition of 'steady state' that would reflect a relative baseline of Interaction/Incident activity as distinct as possible from periods of perturbation. The steady state was defined at the SC level given what we had observed previously regarding the differentiation of impact patterns on different SCs from the same Change.

Each Service component was examined as an hourly timeline based on working hours only. Nine-hour groupings (based on the length of a workday) were examined progressing hourly through the time series. A correction factor was included to allow for long-term trends in the number of impact events over the six-month dataset. A nine-hour segment was deemed to be in steady state, when all values within the period are within the range given by Equation 1.

$$Median(i_t..i_{t+8}) \pm [Standard\,Deviation(i_0..i_{end}) + a] \quad (1)$$

Where i is the number of impact events per hour from the beginning of the data set (i_0) through to the end (i_{end}), t is the hour bin, and a is the long-term trend adjustment. Using this methodology, four distinct periods of activity, as applicable, were determined for each Change-SC pair:

a. SS_{bef}: Period of Steady State (SS) before change implementation for cases that exhibit SS within two working hours leading up to t_0; this period, if uninterrupted, can extend past t_0
b. SS_{aft}: Period of SS after change implementation for cases that exhibit a period of SS up to 27 working hours after t_0, distinct from the SS_{bef} period
c. NSS_{aft}: Period of Non Steady State (NSS) at or after Change implementation; can be followed by a return to SS or proceed uninterrupted for up to 27 working hours

d. NSS_{bef}: Period of NSS up to 27 working hours before change implementation for cases that do not exhibit a period of SS within 2 working hours leading up to t_0

We calculated the mean number of impact events opened for the time periods applicable to each Change-SC pair in both our filtered datasets. The maximum number of impact events opened per hour was recorded for Period NSS_{aft}.

3.2 Decision Tree

Using these parameters, we developed a mutually exclusive, comprehensively exhaustive list of impact patterns, detailed in Fig. 3.

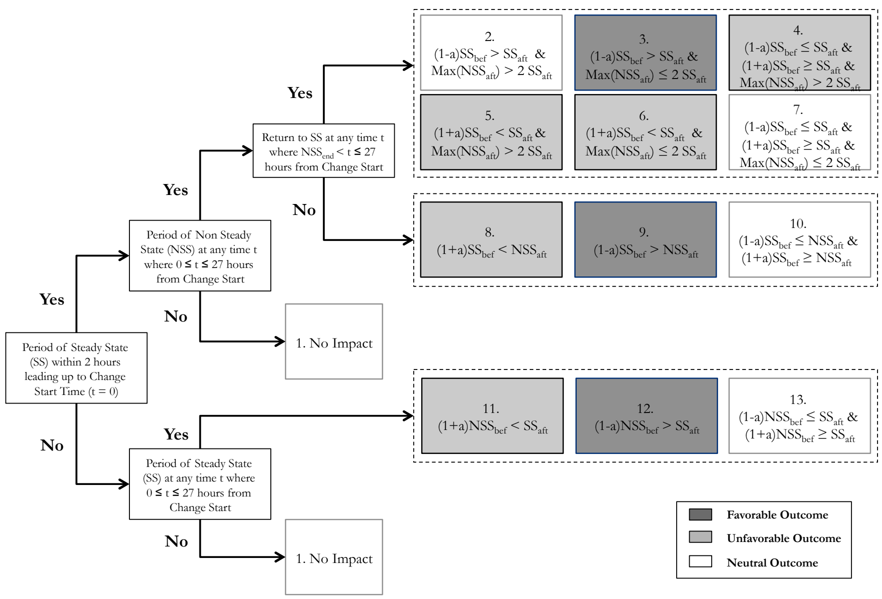

Fig. 3. Impact pattern classification decision tree where a = 0.5 and all values represent the mean number of impact events per hour except where indicated as the maximum number of impact events per hour.

A buffer of 50% was added to the antecedent mean values to ensure that the change in activity levels was meaningful. Similarly, the bar for maximum impact was set at 200% so as to capture a notable spike in workload.

Patterns 2-7 represent cases that exhibit a period of Steady State before change implementation (SS_{bef}), followed by a period of Non Steady State at or after t_0 (NSS_{aft}) with an eventual return to SS within 27 working hours of t_0 (SS_{aft}). Of these, patterns 2 and 3 represent outcomes where SS_{aft} is lower than the SS_{bef}. However, pattern 2 indicates an acute impact spike offsetting this benefit, thus we have classified it as neutral. Patterns 5 and 6 are unfavorable, indicating an increase in Steady State levels of activity, with pattern 5 being less desirable than pattern 6 due to an acute spike during NSS_{aft}. Patterns 4 and 7 reflect cases where no meaningful change is observed in Steady States before and after change implementation. However, pattern 4 is classified as unfavorable given the workload spike during NSS_{aft}.

Patterns 8, 9 and 10 represent cases that exhibit a period of SS_{bef} but not a period of SS_{aft}. These are not as well-defined as patterns 2-7, but we have sought to assign favorability based upon the magnitude of the extended NSS_{aft} period, indicating the directionality of the trend. Pattern 9, which reflects a lower mean of activity levels during the 27 working hours following change implementation compared to the mean of activity levels during the period of SS_{bef}, is desirable, while pattern 8, which captures the obverse, is undesirable. Pattern 10 refers to cases where no meaningful change in activity levels is observed. Patterns 11, 12 and 13 reverse this logic, representing cases that exhibit a period of SS_{aft} but not a period of SS_{bef}, with pattern 12, 11, and 13 being desirable, undesirable and neutral respectively.

Cases that exhibit an uninterrupted period of SS_{bef} continuing through t_0 all the way to 27 working hours after t_0 are classified as 'No Impact' (pattern 1). Conversely, cases that did not exhibit a period of SS at all, either before or after Change implementation were also classified as 'No Impact' due to the lack of any meaningful baseline form which to measure.

Using the decision tree outlined in Fig. 3 we were able to assign all Change-SC pairs to one of the 13 classifications. The pool of isolated Change-SCs was too small to enable effective modeling, so we extended the assignment to all Change-SCs with Start times in scope. In doing so it was necessary to leverage all impact activity observed within the designated time window and SC so as to ascertain an impact pattern, regardless of overlapping Changes that could potentially take place simultaneously.

3.3 Pattern Distributions

Filtering the data for Change-SC pairs that exhibited at least four impact events opened within 27 working hours either before or after Change start reduced the number of Change-SCs in scope from a total of 17,172 over our time frame, to 8,766 and 5,833 for Interactions and Incidents respectively. The impact pattern distribution for the Change-SC pairs in scope across the SD and ITO is given in Table 2.

Table 2. Change-SC volume by pattern type.

		Interactions	Incidents
Pattern	**Outcome**	**Volume**	**Volume**
1	Neutral	6,312 (72.0%)	2,572 (44.1%)
4	Unfavorable	665 (7.6%)	616 (10.6%)
13	Neutral	351 (4.0%)	340 (5.8%)
6	Unfavorable	297 (3.4%)	235 (4.0%)
7	Neutral	272 (3.1%)	85 (1.5%)
2	Neutral	217 (2.5%)	472 (8.1%)
5	Unfavorable	208 (2.4%)	431 (7.4%)
8	Unfavorable	157 (1.8%)	208 (3.6%)
10	Neutral	148 (1.7%)	104 (1.8%)
12	Favorable	39 (0.4%)	234 (4.0%)
11	Unfavorable	38 (0.4%)	171 (2.9%)
9	Favorable	38 (0.4%)	130 (2.2%)
3	Favorable	24 (0.3%)	235 (4.0%)
Total	**N/A**	**8,766**	**5,833**

The majority of cases at the SD (72%) seem to have no impact on workload; the impact of a further 11% is discernable, but not large enough to signify as meaningful change. Nearly 16% of the cases however, corresponding to over 1,300 changes over our time period, exhibit unfavorable outcomes. Within this, patterns 6 and 8 indicate a sustained increase in activity levels following change implementation; pattern 4 suggests a marked surge in workload, while pattern 5 is indicative of both these trends. Favorable outcomes are only evident in a small percentage of cases.

Similar trends are visible at the ITO; however, they are less marked. 44% of cases exhibit no impact, with the impact of a further 17% being neutral. 472 cases over our timeframe, representing 8% of the total volume, fall into pattern 2, which reflects situations in which there is a sustained reduction in activity levels, but only after a surge in workload following change implementation. This can likely be improved through better change planning and execution. 28.5% of the cases exhibit unfavorable outcomes, with the majority of these being constituted by patterns 4 and 5, which also indicate a spike in activity levels at or just after change implementation. 10% of all cases at the ITO exhibit favorable outcomes with activity post-change implementation being lower than levels before without an intermediate period of heightened workload.

We also analyzed these pattern distributions using impact event close times to determine the better method of understanding incoming work. The proportions of patterns 5, 2 and 4 went up, while those of patterns 3, 6 and 7 went down indicating heightened activity on or after change implementation.

3.4 Parameter for Each Impact Pattern

In order to gain a deeper understanding of the impact each of these patterns have on workload, as well the efficacy of constituent Changes, we developed parameters that measure a) the shift in activity levels once a Change is implemented on a SC and b) the average time required to return to 'Steady State' post change implementation.

Activity Levels Table 3 illustrates mean Interaction volumes at the SD during periods of Steady and Non Steady State before and after Change implementation.

Table 3. Activity levels at the Service Desk during periods of interest by pattern (Mean Number of Interactions per Change-SC per Working Hour).

			Mean ± SD				Absolute Change	
Pattern	**Outcome**	**Volume**	**SS_{bef}**	**SS_{aft}**	**NSS_{bef}**	**NSS_{aft}**	**SS,SS**	**NSS,SS**
1	Neutral	6,312	0.6 ± 1.4		24.2 ± 16.9	27.1 ± 13.9		
4	Unfavorable	665	1.1 ± 2.0	0.9 ± 1.7		2.6 ± 4.0	-0.3	
13	Neutral	351		6.5 ± 9.8	9.7 ± 12.6	10.1 ± 12.2		-3.2
6	Unfavorable	297	0.3 ± 0.6	0.7 ± 0.9		0.6 ± 1.0	0.4	
7	Neutral	272	4.8 ± 8.1	6.8 ± 9.4		8.3 ± 10.1	2	
2	Neutral	217	0.4 ± 0.8	0.1 ± 0.4		1.0 ± 1.5	-0.3	
5	Unfavorable	208	0.4 ± 0.9	0.8 ± 1.3		1.8 ± 2.5	0.4	
8	Unfavorable	157	1.3 ± 3.8			7.2 ± 12.5		5.9
10	Neutral	148	8.3 ± 10.5			15.1 ± 14.6		6.8
12	Favorable	39		0.5 ± 1.2	1.6 ± 3.1	1 ± 1.3		-1.1
11	Unfavorable	38		2.2 ± 6.2	8.5 ± 13.4	3.3 ± 4.6		-6.3
9	Favorable	38	0.4 ± 0.7			0.0 ± 0.1		-0.3
3	Favorable	24	0.2 ± 0.4	0.1 ± 0.3		0.0 ± 0.0	-0.1	
Total	**N/A**	**8,766**	**0.8 ± 2.4**	**2.6 ± 6.2**	**14.8 ± 16.1**	**15.5 ± 14.0**	**1.9**	**0.7**

In aggregate, we observe that the average number of Interactions opened at the SD per working hour goes up every time a Change is implemented. The absolute value of this increase stands at two Interactions per working hour per Change for cases in which Steady States are observed both before and after the change, and at 0.7 when there is a transition from a period of Non Steady State to Steady State and vice versa. There is however, a high degree of variance between the different patterns, demonstrating that we have been able to effectively capture divergent impact behavior through their definition. The data suggest that patterns 5, 6, and 8 show a marked increase in activity levels post-Change implementation, while pattern 2 shows a sizeable relative reduction. More definitive however, is the spike in workload observed in the case of patterns 4, 2 and 5, as given by the NSS_{aft} column.

Similarly, Incident volumes at the ITO show a slight increase (Table 4) in the average number of Incidents opened per working hour for each change implementation, for cases that exhibit periods of Steady State both before and after the change. The impact of changes that transition from periods of Non Steady State to Steady State and vice versa is negligible. Patterns 5, 6 and 8 exhibit an increase in activity levels, while patterns 3, 12 and 9 show a small reduction. A sharp surge in workload during the intermediate period of Non Steady State immediately following Change implementation is visible for patterns 4, 2 and 5.

Table 4. Activity levels at the Service Desk during periods of interest by pattern (Mean Number of Incidents per Change-SC per Working Hour).

			Mean ± SD				Absolute Change	
Pattern	**Outcome**	**Volume**	**SS_{bef}**	**SS_{aft}**	**NSS_{bef}**	**NSS_{aft}**	**SS,SS**	**NSS,SS**
1	Neutral	2,572	0.2 ± 0.8		8.1 ± 7.6	4.8 ± 6.2		
4	Unfavorable	616	0.4 ± 1.3	0.4 ± 1.1		0.9 ± 1.9	0.0	
2	Neutral	472	0.2 ± 0.7	0.1 ± 0.3		0.6 ± 1.2	-0.2	
5	Unfavorable	431	0.1 ± 0.5	0.3 ± 0.7		0.7 ± 1.1	0.2	
13	Neutral	340		2.1 ± 4.0	2.8 ± 4.9	2.5 ± 4.1		-0.7
3	Favorable	235	0.2 ± 0.4	0 ± 0.2		0.0 ± 0.0	-0.1	
6	Unfavorable	235	0.1 ± 0.4	0.3 ± 0.6		0.1 ± 0.5	0.2	
12	Favorable	234		0.1 ± 0.4	0.5 ± 1.2	0.5 ± 1.2		-0.4
8	Unfavorable	208	0.2 ± 0.5			0.9 ± 1.6		0.7
11	Unfavorable	171		0.3 ± 0.6	0.2 ± 0.5	0.5 ± 1.3		0.1
9	Favorable	130	0.2 ± 0.5			0 ± 0.2		-0.2
10	Neutral	104	1.9 ± 3.8			4 ± 5.2		2.1
7	Neutral	85	1.7 ± 3.8	0.9 ± 2.7		2.2 ± 4.3	-0.8	
Total	**N/A**	**5,833**	**0.2 ± 1.0**	**0.5 ± 1.8**	**4.0 ± 6.3**	**1.9 ± 3.7**	**0.3**	**0.0**

Impact Duration Table 5 represents the average number of working hours required to return to a period of Steady State following Change Start (t_0, SS_{aft}). In many cases, periods of Steady State that begin before Change implementation extend past t_0. Calculations in Column NSS_{aft} exclude these periods, so as to capture the duration of periods of Non Steady State following Change implementation up to the point that Steady State is achieved. The results are presented by impact pattern for both Interactions and Incidents, and have been limited to cases where a post-Change Steady State is observed.

Table 5. Average number of working hours required to return to Steady State by impact pattern.

		Interactions			Incidents		
			Mean Duration ± SD			Mean Duration ± SD	
Pattern	**Outcome**	**Volume**	**t_0, SS_{aft}**	**NSS_{aft}**	**Volume**	**t_0, SS_{aft}**	**NSS_{aft}**
4	Unfavorable	665	13.1 ± 6.5	5.2 ± 3.9	616	13.7 ± 6.2	7.1 ± 5.5
13	Neutral	351	8.6 ± 7.1	7.3 ± 6.3	340	16.9 ± 7.2	8.0 ± 6.8
6	Unfavorable	297	14 ± 7.8	3.1 ± 1.1	235	13.4 ± 6.2	5.1 ± 4.4
7	Neutral	272	12.8 ± 6.7	5.2 ± 3.9	85	8.8 ± 7.2	5.1 ± 4.0
2	Neutral	217	18.6 ± 7.2	3.8 ± 2.8	472	16.7 ± 7.2	5.0 ± 3.7
5	Unfavorable	208	15 ± 6.5	5.1 ± 3.4	431	10.9 ± 6.5	5.1 ± 4.1
12	Favorable	39	9.1 ± 8.7	8.2 ± 7.9	234	8.6 ± 7.1	8.2 ± 6.8
11	Unfavorable	38	8.8 ± 8.1	5.4 ± 6.9	171	7.5 ± 6.4	7.0 ± 5.8
3	Favorable	24	19.6 ± 10.2	2.8 ±0.4	235	10.7 ± 6.6	5.5 ± 3.4
Total	**N/A**	**2,111**	**13.1 ± 7.6**	**5.0 ± 4.3**	**2,819**	**12.7 ± 7.4**	**6.2 ± 5.2**

In aggregate, it takes the SD and ITO 13 working hours following the start of Change implementation to return to Steady State, with the duration of periods of NSS_{aft} being much shorter. These periods of Non Steady State last longer at the ITO than at the SD, indicating a difference in the handling of workload fluctuations. Breaking down the average duration of periods of NSS_{aft} by impact pattern reveals that cases exhibiting periods of Steady State before the Change (patterns 2-7) return to Steady State sooner than do Changes which transition from a prolonged period of Non Steady State before to a period of Steady State after (patterns 11, 12 and 13). Additionally, one would expect the mean durations of NSS_{aft} for patterns 3 and 6 to be lower than those for patterns 2 and 5 respectively, given that the latter exhibit a spike in activity levels following Change implementation. This trend is evident at SD, but not at the ITO, indicating that Steady State can be achieved at the SD much quicker if the immediate impact of Change implementation is better controlled.

4 Drivers of Impact Pattern

4.1 Multinomial Logistic Regression Model

A multinomial logistic regression was used to predict impact pattern categorizations using Change characteristics. As such, we were able to determine the probability of a particular Change resulting in a given impact pattern. Analyses were performed separately on the Interaction and Incident datasets to differentiate between impact at the SD and ITO.

Selection of Variables We created a list of possible predictors based on the Change characteristics available within the BPIC dataset. These are presented in Table 6 below.

Table 6. Potential Explanatory variables.

Explanatory variable	Definition
Change type (five types)	Change Component, Master Change, Master Change Roadmap, Release, Standard Activity, Standard Change
Origin of the Change	Problem or Incident/Interaction
Downtime	Binary value for whether scheduled downtime was associated with the Change
Weekend	Whether the change occurred on the weekend (defined as 17:00 Friday evening to 08:00 Monday morning)
CI Rollup	Defined as the aggregation of CI Type and Subtype into five distinct categories: Software, Hardware, Network, Personal devices, and Unknown)
CAB Approval	Binary value for whether the Change require approval from the Change Approval Board (CAB)
Risk Assessment	Major Business Change, Business Change, or Minor Business Change

To determine which predictors to use in our analysis, we compared the statistical significance of each independent variable when put in the multinomial logistic regression individually. Furthermore, since the model assumes minimal collinearity, we tested the collinearity between independent variables and excluded variables that were strongly collinear with other explanatory variables. CAB Approval was excluded for this reason. Our final explanatory variables were Change Type, Origin, Downtime, Weekend, CI Rollup, and Risk Assessment.

Other Considerations To avoid over fitting we divided the Change-SC pairs into a training set and a testing set. Change-SC pairs that met the four-impact event threshold were grouped by day within the timeframe. 80% of the days were randomly selected as the training set. The remaining 20% were used as the test set. Multinomial logistic regression and prediction of pattern probabilities was carried out using the R package *mlogit* and the *predict* function respectively.

4.2 Significant Predictors of Impact Pattern

As mentioned previously, our final model for pattern predictions for Interactions and Incidents included whether the change occurred on the weekend, CI rollup, origin of the problem, change type, whether or not there was downtime, and the risk assessment. The output for the Incidents and Interactions are listed in Tables 7 and 8 respectively. The following sections include some notable conclusions regarding this output.

Table 7. Multinomial logistic regression results for Incidents; stars indicated the following significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

	Pattern estimates												
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
intercept		.639	-18.6	-21.1	.352	-19.3	-21.5	-2.55*	-22.6	-21.9	-22.4	-22.8	-21.7
weekend		.193	.186	.6****	.8****	.311	.052	.61***	2.3****	-1.5**	-.701	-.005	.040
ci_rollup:													
network		-.092	-.3**	-.5***	-.073	.588****	1.02***	-1.78	-.69**	-.61	1.3****	1.7****	.76****
none		2.30****	1.50*	.261	-18.9	-18.2	-16.2	-17.8	-18.1	-16.2	-16.9	-16.8	2.30
personal		-19.5	-19.3	3.3****	1.07	-18.8	-17.0	1.45	-18.1	3.3****	-17.4	-17.5	2.9****
software		-.011	-1****	.43****	-.001	-.454**	1.3****	.96****	-.50**	1.8****	.713***	1.8****	1.6****
origin:													
interorinc		2.32	2.71	25.9	2.25	3.39	4.21	2.19	4.36	2.88	3.95	2.24	2.03
problem		.153	-.73**	-.113	-.012	.049	.095	-.149	.195	.313	.430	-.109	.334
change_type:													
master chg		-22.4	-1.98	18.5	-2.04*	17.7	-1.43	-21.0	17.7	-1.60	-1.69	-1.71	17.9
release		-2.3***	16.8	19.3	-2.5***	16.8	19.3	-2.53	16.9	17.3	17.6	18.4	18.6
std activity		-1.76*	16.7	19.3	-1.94**	17.8	18.0	-.894	18.4	17.6	18.2	18.6	19.1
std change		-2.79***	16.7	18.8	-2.7***	16.8	16.7	-1.34	17.2	17.0	16.2	17.7	18.3
downtime		-.032	-.676	.0839	-.271	-.136	-1.52	-.113	-.651	.609	.518	-.777	-.276
risk assess:													
major		.822	1.08	-.151	-19.8	-19.2	-19.3	-19.2	-17.0	-19.1	3.04***	-18.6	-20.6
minor		-.157	.239	.467**	.168	-.461	-.787**	.748**	1.55**	.188	1.49**	1.08***	-3.78**

Table 8. Multinomial logistic regression results for Interactions; stars indicated the following significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

	Pattern estimates												
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
intercept		-20.3	-21.1	1.62	-21.8	-21.1	-48.6	-.486	-37.5	-22.8	-39.4	-22.9	.523
weekend		.208	-17.2	.54****	.98****	1.1****	-.215	.410	1.01**	-.290	-1.03	-.765	-.428**
ci_rollup:													
network		-2.2****	-20.1	-.200	.299	-5.0****	-2****	-.076	-24.3	1.17**	-1.85*	-24.1	1.0****
none		.494	-18.0	.645	-23.6	-.727	-23.6	.593	-20.7	-20.9	-20.9	-20.4	-22.8
personal		-19.3	-14.3	3.6****	2.14*	-19.5	1.52	-18.5	-1.63	5.1****	-16.6	-15.4	5.5****
software		-.291*	1.78**	.026	.59***	-1.6****	.324**	.251	-2****	2.7****	.333	2.1****	1.8****
origin:													
problem		-.342	-1.44**	-.394**	-.202	-.018	.797*	-.034	17.6	.154	20.4	-.035	.595*
change_type:													
master chg		-3.66	-4.19	-3.36**	19.6	18.8	-2.40	-25.9	-1.76	18.5	-2.61	-3.61	-2.97**
release		16.9	16.4	-3.61	17.9	17.7	-2.85	-4.25	16.9	17.1	14.4	-5.67	-4.86
std activity		18.0	14.0	-3.6***	18.5	18.6	-2.0***	-4.02	33.2	18.4	36.0	17.2	-3.4***
std change		18.2	14.9	-3.4***	18.8	17.5	-2.94**	-4****	14.1	17.6	15.3	16.4	-4****
downtime		-.930	-16.5	.336	.245	.170	-1.17	-.335	.204	-1.55	1.78***	.459	.288
risk assess:													
major		-21.0	-16.9	.067	1.51*	-20.4	-19.3	-20.9	-18.3	-21.4	-21.0	3.57***	-20.7
minor		-.608**	.376	-.198	-.63***	.597	-.507**	-.375	.230	-1****	-1****	-.085	-1****

Weekend Changes occurring on the weekend resulted in a statistically significant increase in the likelihood of patterns 4 and 5 for the Interaction model. Coefficients for these relationships range between 0.5 and 1. Pattern 4 represents cases in which there is no long-term change in workload. However, the Non Steady State impact is greater than the second Steady State level, meaning there was a spike in workload. This unfavorable outcome demonstrates that a change occurring on the weekend results in a higher likelihood of increased workload after the change. Pattern 5 also presents an unfavorable workload pattern resulting in a higher level of Steady State activity following the Change compared to the Steady State before the Change, as well as a spike in workload activity during the Non Steady State period.

The Interaction model also demonstrates a strong positive association between a Change occurring on the weekend and pattern 6. Outcome pattern 6 is similar to pattern 5 in that it results in a higher Steady State but it does not have a spike in workload during the Non Steady State period.

This result suggests that there is an unfavorable impact on workload when a Change occurs on the weekend. This, however, may be caused by omitted variable bias since managers may reserve more complex Changes for non-working days.

CI Rollup Configuration Item Rollup is a significant factor associated with five different patterns. When the CI Rollup of the change is 'Network', there is a statistically significant positive association with patterns 6, 7, 11, 12, and 13 for Incidents. Patterns 6 and 11 are unfavorable outcomes, 7 and 13 are neutral outcomes, and 12 is a favorable outcome. The model for Interactions, however, demonstrates a statistically significant negative relationship with patterns 2, 6, and 7 as well as positive relationships with patterns 10 and 13. While 2, 7, 10, and 13 are all neutral outcomes, pattern 6 is an unfavorable outcome.

CI Rollup of 'None' has no statistically significant relationships in the Interaction model. In the Incident model, 'None' has a highly statistically significant positive relationship with the neutral pattern 2. The effect is also substantial, with a coefficient of 2.30. This suggests that a less specific Change has a minimal impact on the workload after the change and while the change occurs.

CI Rollup of 'Personal' has a highly significant positive relationship with patterns 4, 10, and 13 for both Incident and Interaction models with coefficients between 1 and 6. Pattern 4 results in increased workload in the short run but no difference in Steady State. Pattern 10 represents when the workload level is in a Steady State before the Change start but does not result in a Steady State after the Change. Pattern 13 exhibits the opposite where there is a Steady State only after the Change. This follows the logic that the Change is addressing a personal issue since it would result in minimal impact on the long-term Steady State in all cases, but might increase workload in the short run while this personal issue is addressed.

CI Rollup 'Software' has statistically significant associations with all patterns between the two models. Since this represents the majority of patterns with conflicting favorability, there is little we can extrapolate from this outcome.

Origin The Origin of the Change has little significance on the pattern outcome with a few exceptions. In the Incident and Interaction models, there is a negative relationship with the origin being a 'Problem' and pattern 3. Pattern 3 is a favorable outcome in which there is a long-term reduction in the workload after the change as well as minimal impact during the Change. This could be due to the fact that a Change originating from a problem could have large effects on workload and might therefore result in an influx in work right after the Change. In the Interaction model, there is also a negative statistically significant relationship between 'Problem' and pattern 4. Pattern 4 represents a negative outcome in which there is no Change in steady state but there is an increase in workload after the Change open during the Non Steady State period. This could be due to the fact that Changes resulting from Problems might be less likely to cause a short burst of interactions because they are more serious issues.

Change Type A Change Type of 'Master Change' is not a significantly associated with any pattern for Incidents, but has statistically significant negative relationships with patterns 4 and 13 in the Interaction model. Both patterns do not reflect long term increases in workload for the service desk. Since a 'Master Change Type' suggests that it is larger scale, it makes sense that these pattern outcomes would be less likely because the Change would probably have a larger scale impact on the influx of calls to the service desk.

'Release' has a positive statistically significant relationship with pattern 2 for Incidents, but no strong correlations with any patterns in the Interaction model. Pattern 2 is a neutral outcome in which the Steady State does not change dramatically after the change and there is not a large influx of workflow between the Steady States. This demonstrates that there is an association between this Change Type and minimal impact on workload. This Change Type also has a negative statistically significant relationship with pattern 5. Pattern 5 is unfavorable, since it results in a higher long-term Steady State of workload and an influx of work during the non-Steady State period. Since Release has a negative correlation with this pattern, it is less likely that this workload pattern will result under this condition. Since the Change type of 'Release' suggests that there is an introduction of a new technology version, it follows that there will be a correlation with incidents opened due to SD agents that will not be as familiar with the technology. Therefore while there may not be a large change in interactions, the proportion of those interactions that will be escalated to incidents may rise, potentially resulting in a more significant relationship in the Incident model.

'Standard Activity' has a minimal statistically significant relationship with Incident pattern outcomes. It, however, has negative associations with patterns 4, 7, and 13 in the interactions multinomial logistic model. All of these pattern outcomes are neutral. While specific pattern outcomes are not necessarily more likely when there is 'Standard Activity', it is less likely that a neutral outcome will occur.

In the interaction model, 'Standard Change' has a negative statistically significant relationship with the same three patterns (4, 7, and 13) as well as pattern 8. Pattern 8 represents when there is no steady state before the change but a steady state higher than the Non Steady State average. This shows that there is less likely to be no impact on workload as well as less likely to be stabilization in combination with a long-term increase. On the other hand, 'Standard Change' has a negative statistically significant relationship with both pattern 2 and pattern 5 in the Incident model. While pattern 2 is neutral, pattern 5 has the disadvantage of long-term increase in workload as well as a high influx of work between the Steady States. This can be interpreted that while a Standard Change does not have minimal impact on workload, it also does not result in the worst-case scenario workload.

Downtime Whether or not downtime is associated with the Change is not a significant factor affecting workload pattern outcome for Incidents. However, in the Interaction model, there is a statistically significant positive relationship with pattern 11 with a coefficient of 1.78. Pattern 11 represents an outcome where there is not a Steady State before the Change but results in a high Steady State after the change. This increase in Interactions after the Change could be due to the fact that downtime could occur for a period of time after the Change. This relationship could be picking up the influx of service desk calls during the downtime for those asking why a service is not working. Because these phone calls are easily resolved, it follows that there would not be an impact on Incidents.

Risk Assessment Risk Assessment is a significant factor affecting workload outcome patterns. A 'Major Business Change' has a positive statistically significant association with pattern 11 in both models. Pattern 11 represents a pattern where there is not a Steady State before the Change but the average workload level during the Non Steady State period is lower than the Steady State that occurs after the change. This could be due to the fact that a Major Business Change was implemented from an inconsistent influx of work. Additionally, the Steady State in the short term, because it is an important business change, may result in a high level of work since it affects a large part of their business. However, outside of the 27-hour window it is possible that the Change could result in reduced workload.

In the interaction model, a 'Minor Change' has significant negative relationships with patterns 2, 5, 7, 10, 11, and 13. These are all negative or neutral outcomes that are less likely to occur with a minor change. On the other hand, a Minor Change has a statistically significant positive relationship with pattern 12 in the Incident model. Pattern 12 is a favorable outcome in which there is no Steady State before the Change but the Steady State after the Change is lower than the mean of the Non Steady State period before the Change. This could result since a minor Change could be implemented due to this unsteady influx of work. Once this problem is corrected by the Change, since it does not affect as many operations it can quickly be resolved into a lower Steady State.

4.3 Prediction of Daily Impact Patterns

Three separate prediction approaches were implemented and compared for accuracy: predictions based on assuming that the pattern distribution in the training set mirrored the distribution of the test set, predictions based on a Multinomial Logistic Regression model and predictions assuming all changes did not create any impact pattern in the service component. This final prediction approach is used to create a nave baseline that does not anticipate Change impact against which the other models could be compared.

All three of these models were compared for accuracy based on both the open and close times of impact events. The analyses showed a greater accuracy in predictions based on open times. This is expected, as impacts directly linked to Changes are likely to originate soon after the Change is made. Impacts that close due to a change may take some time for the Change to have an effect. This disparity in accuracy led to impact events open times being used as the time point of reference for the prediction comparisons.

The first prediction model investigated was based on simple Change pattern distribution. Changes in the data set were split into a training set and a test set as outlined previously. The distribution of pattern types within the training set was recorded and used as the basis for prediction. The distribution from the learning set was then applied to the test set and its accuracy in comparison to the actual values was recorded by comparing predicted values by SC to the actual values on a daily basis. The number of incorrect predictions (from either not predicting a change that occurred, or by predicting a change that didn't actually occur) was compared with the total number of changes. This created a metric between -1 and 1 with 1 having predictions matching actual patterns exactly and -1 failing to predict all actual patterns while at the same time predicting patterns that did not exist. This was done on a daily basis in order to best reflect the proposed usage of such a model: to predict the number of favorable or unfavorable near-term impact events expected from scheduled Changes on a given day.

With the goal of further improving accuracy, a multinomial logistical model was created from the same learning set as the distribution predictions. The overall accuracies of the predictions of each model were then compared. The results of this comparison for both Interactions and Incidents are shown in Table 9.

Table 9. Accuracy of predicting impact patterns via simple Change distribution, multinomial model, or nave assumption of no impact.

	Interactions			Incidents		
	Model	Distribution	'No Impact'	Model	Distribution	'No Impact'
Avg. Accuracy	0.456	0.459	0.405	0.044	0.029	-0.246
No. days with highest accuracy	13	19	0	16	15	0

The predictions based on the simple distribution of the training data show on average greater accuracy in predicting patterns than the baseline assumption that no quantifiable patterns exist but clearly have their limitations. The predictions based on Incidents show much lower accuracy due to a smaller sample size.

Both models are reasonable predictors of patterns in comparison to the assumption that Changes create no discernible impact. The multinomial model provides more accurate results than the distributed model for Incidents, with less success predicting Interactions. This analyses shows that predictive models can be utilized to anticipate service desk workload with some success.

5 Performance Metrics

5.1 Metric Definitions

Project managers are responsible for a specific set of SCs across all sub-streams of the process flow and are expected to deliver the same or improved service levels after each Change implementation [6], [5]. In order to test if this expectation is being met, we evaluated performance trends using the metrics described below.

Performance at the Service Desk was measured as the percentage of Interactions affecting a particular Service Component that achieved 'First Call Resolution'. This common industry benchmark measures the capability of the SD to close an Interaction in one instance without requiring escalation. 'Number of Steps to Resolve' an Incident was used as a proxy for Incident complexity. This was calculated using the total number of activities before the Resolved Time per Incident ID in the Incident activity data. We assumed that the Handle Time Hours as given in the BPIC dataset would be an accurate reflection of the quantity of effort expended by the ITO [6], [5].

We calculated these metrics as a function of a) the incremental number of completed Changes, which enabled us to assess manager performance as it relates to Change implementation, and b) a function of time, which provides a measure of overall performance independent of Change implementation. Change end time was used as the indicator that the Change had been fully implemented. Performance metrics were measured as an average across all Interactions and Incidents following each incremental Change. For analysis over time, the First Call Resolution metric was calculated as a percentage across all Interactions that were opened on the same day.

We used a linear regression model to fit a trend line to the data using the least squares approach with a significance threshold of $p \leq 0.05$. The slope was extracted as a measure of performance improvement or deterioration for significant trends. SCs with fewer than three data points were excluded from the analysis, while SCs with no related Changes were excluded from the analysis performed over incremental number of Changes. There are 293 unique SCs in the Interactions records dataset, 279 in the Incident records and 277 in the Change records.

In aggregate, there is a general trend toward improved or constant service levels over time. Overall, the Number of Steps to Resolution decreased by 0.23 steps each month, and mean Handle Time decreased by 2.1 hr each month. The percentage of First Call Resolution remained constant at 64%, translating a constant service level over time, however below the 74% industry benchmark [7].

5.2 Performance Trends by Incremental Change

Fig. 4 illustrates performance change as measured on SC WBS000090. In this case, the First Call Resolution at the SD is improving, while the Handle Time and Steps to Resolution at the ITO are worsening. This suggests an improvement in knowledge transfer to the SD, offset by increasing complexity of Incidents being handled by ITO with each incremental Change, potentially indicating a decline in the quality of Change planning.

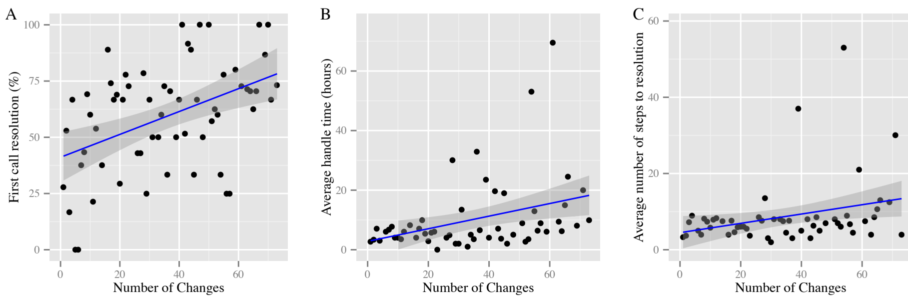

Fig. 4. Service Component WBS000090 performance metrics by incremental Change: **A.** First Call Resolution; **B.** Handle Time; **C.** Steps to Process.

98% of SCs (271 out of 277) show unchanged or improved service levels as a function of incremental Change implementation. The percentage of SCs showing no change over incremental number of changes was 96% for First Call Resolution, 96% for Steps to Resolution and 94% for Handle Time. 26 SCs showed a significant change in service level either at the SD, the ITO, or both, and this accounted for 9% of all SCs. The top and bottom-performing 5 SCs are presented in Table 10; complete results are included in Appendix Table A1.

Table 10. Top and bottom 5 Service Components by performance shift per incremental Change; for First Call Resolution and Time to handle only 2 SCs showed significant performance worsening and for Steps to Resolution only 4 SCs.

First Call Resolution		Time to Handle		Steps to Resolution	
SC WBS000	ΔPerf (%/change)	SC WBS000	ΔPerf (min/change)	SC WBS000	ΔPerf (steps/change)
149	6.59	055	-560	243	-2.14
242	5.65	153	-233	153	-0.857
330	3.3	088	-136	139	-0.291
002	0.96	027	-104	095	-0.025
900	0.48	079	-94	162	-0.019
073	-0.03	090	13	102	0.001
203	-0.23	157	105	090	0.122
				157	0.359
				149	1.236

5.3 Performance Trends Over Time

Studying performance metrics over time allows us to capture changes in performance resulting from overall effort of the project managers to improve service levels without relying on Change implementation. Following this strategy, we found that 91% of SCs succeeded in demonstrating the same or improved service levels over time.

Table 11 shows that there are more SCs that demonstrate a change in performance over time compared to the per change analysis. The majority of these SCs showed improved service level: 68% showed a gain in performance for First Call Resolution, 63% for Steps to Resolution and 74% for Handle Time. 9% of the SCs showed a loss in performance at either level of the process flow. The top and bottom-performing 5 SCs are listed in Table 12; complete results are included in Appendix Table A2.

Table 11. Service Components data availability and trends.

	Metrics by Incremental Change			Metrics by Time		
	First call resolution	Handle Time	Steps To Resolution	First Call Resolution	Handle Time	Steps To Resolution
SCs with ≥ 3 data points	183	157	155	246	221	221
Improved	8	14	7	21	20	19
Worsened	2	2	4	10	7	11

Table 12. Top and bottom 5 Service Components by performance shift over time.

First Call Resolution		Time to Handle		Steps to Resolution	
SC WBS000	ΔPerf (%/day)	SC WBS000	ΔPerf (min/day)	SC WBS000	ΔPerf (steps/day)
079	3.09	055	-147	210	-1.10
220	0.90	329	-130	051	-0.61
059	0.73	162	-82	330	-0.58
126	0.68	088	-69	187	-0.56
042	0.64	330	-67	088	-0.48
062	-0.34	090	5	090	0.26
008	-0.45	070	5	186	0.31
322	-0.67	256	7	149	0.57
144	-0.94	251	13	256	0.58
137	-4.16	157	37	157	0.89

A look at individual SCs corroborated the trends seen over incremental Changes. WBS000088 for instance, showed a reduction of over 2 hrs in Handle Time per Change, which translated to 1 hr of improvement per day, as seen here. WBS000330 did not show any direct reduction in Handle Time or Steps to Resolution performance resulting from Change implementation, but we saw significant reduction of over an hour in Handle Time per day, and 0.48 Steps to Resolution per day, when viewed holistically.

These two approaches therefore complemented each other and enabled us to measure service level trends either across time or as Changes were implemented.

However, it did not allow us to compare general performances of project managers over the time period studied. To this end, average performance metrics were developed and used.

5.4 Manager Performance Dashboard

We developed a prototypical dashboard enabling a rapid assessment of manager performance across the two dimensions of the SD and ITO. To this aim, system-driven standard times to resolution were developed for Incident resolution as a benchmark for each Incident Type-Subtype pair. This was accomplished by measuring the working hours[4] between the Open Time and Resolved Time for each Incident and taking the statistical median across all Incidents of the same Type-Subtype pair. Out approach assumed that the standard times would be similar for Incident Type-Subtype pairs across SCs. Appendix Table A3 summarizes the calculated benchmark time for the 71 Type-Subtype categories.

For each SC, performance at the SD was visualized on the dashboard as the percentage of First Call Resolution. This was plotted against the proportion of Incidents resolved within the standard time. Fig. 5 presents the proof-of-concept manager performance dashboard enabling simplified performance monitoring across all SCs and both sub-streams. The number of working hours to resolve an Incident accounts for both the time spent working on an Incident and the time lost in between two activities.

Each dot was weighted by the total number of working hours spent on each SC, highlighting SCs that resulted into the highest workload at the ITO. SCs with a 100% First Call Resolution performance were not represented on this diagram as such SCs never resulted in an Incident.

The diagram showed that there was no systematic relationship between the performances at the SD and ITO, and no obvious trend in performance across all SCs. The percentage of First Call Resolution ranged from 0% to 100% with an average at 64%. 64 SCs (22%) performed better than the industry benchmark (74%), highlighting significant room for improvement for the remaining 78%.

At the ITO, the percentage of Incidents meeting expectations also ranged from 0% to 100% depending on SC. 49% of SCs showed a performance of 50% and higher, which also highlighted significant room for improvement. Finally the size of the dots enabled us to observe that six SCs were accounting for a large proportion of the total workload. WBS000073, WBS000072, WBS000263, WBS000162, WBS000088, WBS000091 used 53% of the ITO total working hours. Improving the management of these SCs emerges as an efficient way to significantly reduce the workload at the ITO.

[4] The function used to account for office working hours assumes that operators are working from 8am to 5pm, Monday to Friday. Official holidays were also considered as off. Those assumptions were supported by analyzing the time/day stamps logged into the provided records.

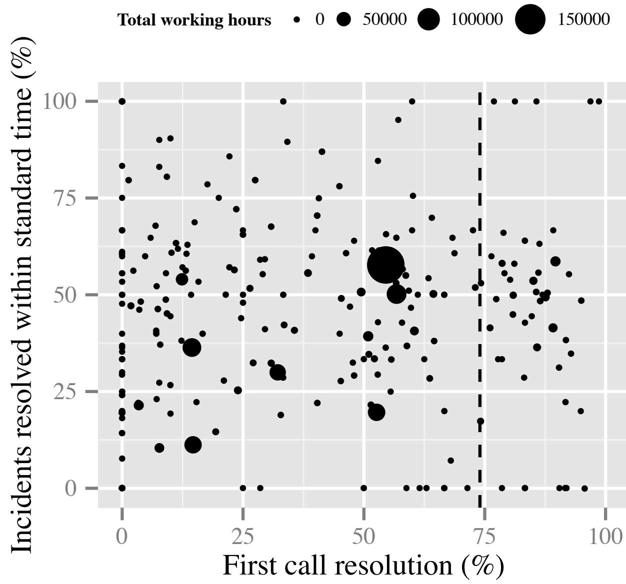

Fig. 5. Product manager performance dashboard; Each Service Component is plotted by the proportion of First Call Resolution (includes industry benchmark of 74%) vs. the proportion of Incidents resolved within the standard time.

6 Incident Resolution Process Improvements

During our investigation, we found that a large proportion of the ITO workload was generated by a relatively small subset of Incidents. By identifying and remediating process bottlenecks for these Incidents, we could significantly improve the ITO process and reduce resource expenditure.

In order to identify patterns in those Incidents, we isolated the top 1% of Incidents having the highest Handle Times. This subset comprised 423 Incidents and accounted for 26% of the total Handle Time. Process data for these Incidents were loaded into Disco process mining software for further analysis. The subset was compared to the overall Incident population across three characteristics: the SCs causing the Incidents, the teams handling the Incidents, and the types of activities involved in their resolution.

6.1 Service Components and Teams of Interest

Ten SCs accounted for 83% of Incidents in the top 1% (see Appendix Table A4). Correspondingly, a small subset of teams are associated with handling these Incidents and SCs (see Appendix Tables A5-A6) Specifically investigating and improving the process flow for these SCs and teams could significantly decrease the total workload at the ITO.

6.2 Activity Types

When comparing the frequency of activities between all Incidents and top 1%, we observed a strong enrichment in 'Assignment', 'Reassignment', 'Operator Update', 'Update', and 'Update from customer' activities. Assignment/Reassignment activities comprised 43% of top Incident activities compared to 30% overall. This suggests that either these Incidents were not assigned appropriately or that these complex Incidents required the involvement of multiple teams to resolve. The complete breakdown of activity types is given in Appendix Table A7.

6.3 Process Bottlenecks

Analysis of the process map for these Incidents showed increased time spent between two specific pairs of activities: 'Update from customer' followed by 'Closed' and 'Update' followed by 'Closed'. Out of 423 cases, 58 took a median of 54.6 days to complete the first path in one step and 62 took a median of 5 days to complete the second path in one step. This compares with 14 min and 53 min respectively for the overall Incident population. Although this appears to be time spent waiting for more information, it is correlated with increased Handle Time. There are two potential explanations: either there is significant effort being expended fruitlessly during this waiting period, or Handle Time does not accurately reflect the amount of effort expended in handling an Incident.

Additionally, the frequency of 'Assignment', 'Reassignment', 'Update', and 'Operator update' showed a 5-fold increase for top Incidents. This could partially be explained by the assumption that these are highly complex Incidents. However, it may indicate a sub-optimal Incident assignment process.

To illustrate these findings, the process flows of the SCs that were highly enriched in the top 1% Incidents were studied. Three different patterns were observed.

Inefficient Incident Assignment WBS000072s and WBS000263s slow process flow seem to be due to a lack of appropriate Incident routing. The frequency of 'Assignment', 'Reassignment', 'Update', and 'Operator update' activities drastically increased to eight per case, with a concomitant increase in the median time to complete Update activities ranging from a few hours to several days. Consequently, back and forth between operators and slow update processes seem to be the major source of inefficiency in the process flow.

Slow Customer Updates WBS000088 and WBS000162 showed similar process flow degradation that could be explained by two mechanisms. On one hand, the frequency of Assignment and Reassignment activities were strongly increased up to six per case. In addition, 85% of the Incidents went through an 'Update from customer' phase that took a median of 50 days to be completed.

WBS000055s and WBS000167s process flow degradation could also be explained by the same mechanisms, but with different weight in the final outcome. WBS000167 slow process was mostly due to an increase in the number of Incident transfers, as evidenced by the strong increase in the frequency of Assignments (11/case), Reassignments (7/case), Operator updates (8/case), Update (6/case) and Status change (5/case). On the other hand, WBS000055 worsening of process flow was mostly due to slow customer updates, involving 82% of the Incidents with a median of 46 days to completion.

Dependency on External Contractors For WBS000073, several paths involving external entities were considerably slowed down, such as the paths involving 'External vendor assignment', 'Communication with vendor', 'Update from customer', and 'Communication with customer'. Considering that 75% of the cases were assigned to an external vendor and 50% involved communication with a customer, WBS000073 was strongly dependent on external entities activities. Additionally, an increase in Assignments (6/case) and Reassignments (4/case) frequency also participated in the process degradation.

Incident Resolution Delay The Incident activity records enabled us to monitor the elapse time necessary to complete specific activities for the top 1% Incidents and to compare it to a system-derived standard time. Using Disco, we identified the steps in the process accumulating high amount of elapsed time. Specifically, 'Update from customer' followed by 'Closed' and 'Update' followed by 'Closed' respectively reached a total of 87.5 months and 63.8 months. Using the median of time required to complete these steps across all Incidents as the standard time, which are respectively 14.2 min and 53 min, we estimated a loss of about 64,000 hours and 56,500 hours at each respective bottleneck. These times can not be directly translated into actual handle time and seem to account primarily for 'waiting time'. Consequently, being able to accurately monitor the actual working time at the activity level would enable us to more efficiently measure the human effort lost at each bottleneck and to provide a better estimate of the potential reduction in total workload at the ITO.

7 Discussion

In order to identify unlinked Interactions and Incidents resulting from the impact of Change activity, we assumed that all Interactions, Incidents and Changes were related within one Service Component within a set time window. This was a key limitation of our analysis due to the paucity of explicitly linked impact events. Since our analysis focused on relative fluctuations in activity levels within particular service offerings, this was a reasonable assumption to make. This approach however, did not take into account the potential for cross-impact i.e., a Change on a particular SC triggering an Incident or an Interaction on another SC. Conversely, events could have been inadvertently linked to unrelated Changes. This noise limits the accuracy of our predictions.

Our impact pattern assignment system produced a distribution of impacts patterns that varied considerably. Application of this distribution to scheduled Change was able to more accurately predict the level of impact at the SD and

ITO than an assumption of no impact, despite the fact that 'No Impact' is the most prevalent largest pattern type that we encountered. The multinomial logit model sought to leverage various Change parameters as predictors of impact, it did not perform consistently better than a simple expected impact distribution. The brevity of the dataset exacerbated this. The number of variables known for each Change was large in comparison with the size of the training set. As such, Change volume associated some patterns in the training set was relatively low.

Based on our experience with similar datasets from prior work, it is common for such explicit linkages to be absent from ITIL process data unless there is an enforced effort to track these relationships. Furthermore, applying text mining techniques to descriptor fields and communications between support personnel often reveals linkages not entered into workflow management systems. Increased tracking of these linkages would help isolate the events that are related to each other while minimizing noise, leading to more accurate measures of impact patterns. This information will allow for monitoring of Change impact on a much broader, more comprehensive scale facilitating more accurate planning and resource allocation. In addition, this data can be used to pre-empt Incidents, both major and minor, by highlighting potential problem areas that need to be planned and tested for before implementation.

Furthermore, a list of Service Components for which project managers are responsible would enable us to analyze the data at a more functional, more appropriate level of abstraction. We would get a better sense of the realms within which a Change can cause downstream activity, allowing us to bring cross-impact Interactions and Incidents into the analysis. This would help create a much more comprehensive and nuanced picture of impact patterns and performance evaluation.

Measuring manager performance by incremental Change allows for the detection of improvement or deterioration at a very granular level in a manner that is directly correlated to Change implementation. Supplementing this metric with measurements over time provides a comprehensive means of tracking product manager performance inclusive of Change implementation, workflow improvements, and resource allocation. Metrics for some Service Components however, could not be measured due to low volumes of Interactions, Incidents or Changes. This indicates low levels of workload across the service offering, which is a good sign overall. Performance for these SCs however, must be measured in light of Interaction and Incident volumes.

Much of our analysis has leveraged the volume of incoming work with the assumption that all work items require an equal amount of effort to address. Given that our focus is on comparing relative activity levels within a particular service offering, this is a fair assumption to make. However, managers must staff according to effort, skill and knowledge, not just work item volume. The system-driven standard times generated above, which indicate the median effort required to address Incident Type-Subtype pairs across each SC, not only allow for a transition to data-driven staffing models, but also define work units at which skill and experience can be evaluated.

As the results above show, much of the delay in Incident resolution is spent in routing tickets to the appropriate team and technician. External vendors need to be called in on a number of cases. A database of in-house skills and experience across teams and geographies would enable much more efficient resource allocation and significantly reduce both external dependency and the time to resolve.

8 Next Steps

Our initial exploration of the BPIC 2014 dataset led us to identify 13 distinct impact patterns of Changes on SD and ITO activity at the SC product level. Using our predictive model, we were able to leverage Change characteristics and historical trends to determine the likelihood of a change having a particular kind of impact, and by extension the distribution of impact patterns within a particular SC over a set time period. While the relative impact of these patterns has been well-defined in terms of changes in mean activity levels and average impact durations, the next step would be to quantify this impact i.e., determine the absolute change in activity levels associated with each pattern. The challenge here will be to build a time series model that looks at historical trends to capture fluctuations in activity levels over time, something that the current brevity of the dataset does not allow. These results can then be combined with the output of the predictive model outlined above to estimate changes in incoming work volume. In practice, product managers would be able to simply feed in the relevant parameters of all scheduled or expected Changes over a particular time period, and get an estimate of the direction and degree to which their activity levels will be affected. Supplementing these volume estimates with knowledge of the amount of effort required to address these work items, managers can staff accordingly. Awareness of the composition of this incoming work, combined with the skills inventory outlined above can be leveraged to optimize shift scheduling and ticket assignment. As such, with a larger dataset and some additional data points, the elements detailed in our analysis can be brought together to in a manner that helps managers optimize service delivery.

From a business perspective, we found that Changes, Interactions, and Incidents have some elements of connection. Based on our actual experience, the initial manipulation of the data files and the provision of these files for BPIC was incomplete to entirely link these events properly. Technicians in the banking industry involved in these processes communicate heavily through the use of unstructured data via comment fields, email, and instant messaging platforms. By performing proper analyses between the structured data elements included in the challenge, data elements that are present in Rabobanks Incident, Change, Problem, and Request management systems and unstructured data that reveal often how the work was performed and how it was linked together, we would be able to ascertain root causes. The same types of analyses would create stronger predictors and allow Rabobank to better manage this area. The potential benefits include reducing the number of Changes that create unnecessary Incidents, highlighting potential weaknesses in testing processes, and identifying resource gaps in certain specialties, leading to a more effective and efficient organization that could be monitored on a real time basis.

References

1. Machackova, E., "Change management practically: Step-by-step approach to implement ITIL compliant process, activities and roles", CreateSpace Independent Publishing Platform (2013)

2. Valverde, R., Sade, R.G., Talla, M., ITIL-based IT service support process reengineering, Intelligent Decision Technologies - IT Service Management and Engineering: An Intelligent Decision-Making Support Systems Approach, vol. 8, pp.111-130, IOS Press Amsterdam (2014)

3. Kowalczuk, Z., Orlowski, C., "Advanced Modeling of Management Processes in Information Technology", Springer Publishing Company (2013)

4. IEEE Task Force on Process Mining. "Business Process Intelligence Challenge (BPIC)." 10th International Workshop on Business Process Intelligence 2014, http://www.win.tue.nl/bpi/2014/challenge

5. Rabobank, Quick Reference Guide. IEEE Task Force on Process Mining (2014), http://www.win.tue.nl/bpi/_media/2014/quick_reference_bpi_challenge_2014.pdf

6. (2014, May 15) Fluxicon Webinar. In BPI Challenge, https://global.gotowebinar.com/join/5632881989969890562/268786484

7. The Pink Elephant IT Management Metrics Benchmark Service: IT Service Management Metrics Benchmarks: 2013 Update, Pink Elephant (2013), http://www.pinkelephant.com/uploadedFiles/Content/en-us/ResourceCenter/PinkPapers/Metrics-Benchmark-White-Paper-2013.pdf

Appendix

Table A1.

First Call Resolution		Time to Handle		Steps to Resolution	
SC WBS000	ΔPerf (%/change)	SC WBS000	ΔPerf (min/change)	SC WBS000	ΔPerf (steps/change)
149	6.59	055	-560	243	-2.140
242	5.65	153	-233	153	-0.857
330	3.30	088	-136	139	-0.291
002	0.96	027	-104	095	-0.025
090	0.48	079	-94	162	-0.019
125	0.41	139	-70	206	-0.007
088	0.33	298	-59	073	-0.005
255	0.15	066	-59		
		162	-32		
		152	-24		
		219	-18		
		268	-16		
		073	-2		
		095	-1		
73	-0.03	090	13	102	0.001
203	-0.23	157	105	90	0.122
				157	0.359
				149	1.236

Table A2.

First Call Resolution		Time to Handle		Steps to Resolution	
SC WBS000	ΔPerf (%/day)	SC WBS000	ΔPerf (min/day)	SC WBS000	ΔPerf (steps/day)
079	3.09	055	-147	210	-1.10
220	0.90	329	-130	051	-0.61
059	0.73	162	-82	330	-0.58
126	0.68	088	-69	187	-0.56
042	0.64	330	-67	088	-0.48
177	0.46	066	-27	138	-0.43
273	0.45	192	-25	151	-0.39
085	0.44	298	-24	066	-0.36
012	0.41	172	-23	206	-0.36
149	0.30	219	-9	298	-0.29
330	0.30	239	-8	162	-0.28
219	0.22	152	-7	239	-0.25
016	0.12	145	-7	294	-0.23
088	0.12	263	-6	152	-0.20
255	0.10	318	-5	219	-0.11
095	0.10	151	-5	318	-0.08
292	0.09	073	-3	095	-0.08
318	0.07	094	-2	183	-0.06
223	0.07	228	-2	073	-0.02
296	0.05	095	-1		
092	0.04				
199	-0.11	092	3	128	0.07
172	-0.13	128	3	072	0.13
025	-0.27	090	5	102	0.16
120	-0.31	070	5	181	0.17
280	-0.33	256	7	271	0.18
062	-0.35	251	13	217	0.23
008	-0.45	157	37	090	0.26
322	-0.67			186	0.31
144	-0.94			149	0.57
137	-4.16			256	0.58
				157	0.89

Table A3. Median handle time and median working hours per type-subtype category.

Type-Subtype Category	Median Working Hours	Median Handle Hours	Number of Incidents
application Server Based Application	5.8	4.0	15,354
application Web Based Application	5.1	4.0	8,806
subapplication Web Based Application	1.9	1.0	6,058
application Desktop Application	8.3	4.0	3,350
software System Software	1.0	0.0	2,291
computer Laptop	23.1	17.0	1,910
subapplication Server Based Application	1.4	1.0	1,524
application SAP	10.4	9.0	1,184
storage SAN	6.6	6.0	1,126
computer Banking Device	7.9	7.0	912
application Citrix	2.6	2.0	532
computer Desktop	22.0	21.0	525
application Client Based Application	10.0	8.0	431
computer Windows Server	11.3	5.0	313
hardware DataCenterEquipment	8.6	0.0	304
application Exchange	2.5	2.0	195
displaydevice Monitor	27.6	21.5	178
database Database	2.3	2.0	164
application Standard Application	14.9	11.5	154
subapplication Citrix	13.7	8.0	149
storage Controller	4.9	0.0	146
hardware MigratieDummy	3.0	2.0	120
no type no subtype	4.2	2.0	92
networkcomponents Switch	2.7	2.0	90
officeelectronics Scanner	0.6	0.0	67
networkcomponents Network Component	5.2	1.5	64
computer Omgeving	3.4	2.5	60
officeelectronics Printer	5.3	4.0	53
networkcomponents Router	2.1	1.0	43
computer Thin Client	12.1	10.0	40
networkcomponents Net Device	0.9	0.0	38
software Automation Software	7.9	5.5	36
computer Linux Server	2.7	0.0	33
computer VDI	8.1	3.0	29
subapplication Standard Application	7.7	6.0	29
hardware KVM Switches	15.7	17.0	28
hardware Keyboard	13.2	0.0	24
hardware Encryption	7.4	2.0	19

Type-Subtype Category	Median Working Hours	Median Handle Hours	Number of Incidents
networkcomponents Lines	1.7	1.0	16
application SharePoint Farm	28.0	7.0	14
computer Oracle Server	10.0	12.5	12
computer X86 Server	10.5	4.0	11
computer ESX Cluster	4.7	4.5	8
applicationcomponent MQ Queue Manager	11.0	8.0	7
computer NonStop Server	4.3	0.0	7
database RAC Service	7.1	2.0	7
database Instance	12.5	2.0	5
networkcomponents Firewall	2.0	2.0	5
subapplication Client Based Application	40.6	12.0	5
computer Unix Server	18.8	10.5	4
storage Switch	12.1	0.5	4
computer zOS Cluster	3.6	6.0	3
networkcomponents IPtelephony	39.3	1.0	3
software Server Based Application	4.1	4.0	3
computer Neoview Server	3.3	7.5	2
computer NonStop Harddisk	2.0	1.0	2
computer zOS Systeem	7.1	1.0	2
software Database Software	16.7	6.5	2
application VMWare	0.7	0.0	1
computer Appliance	26.0	18.0	1
computer ESX Server	33.0	5.0	1
computer Windows Server in extern beheer	7.5	0.0	1
database Virtual Environment	0.0	0.0	1
hardware UPS	84.6	0.0	1
networkcomponents Iptelephony	1.1	1.0	1
Phone Number	61.3	71.0	1
storage Neoview Server	11.3	12.0	1
storage Tape Drive	7.1	8.0	1
storage Tape Library	4.9	0.0	1
storage Virtual Tape Server	8.8	10.0	1
subapplication Exchange	0.9	0.0	1

Table A4. Ten most frequent SCs for top 1% Incidents.

Service component	Relative frequency for top 1% Incidents (%)	Average handle time (hours)
WBS000162	19.39	595
WBS000088	18.68	599
WBS000073	10.64	459
WBS000072	9.46	372
WBS000055	6.86	670
WBS000263	6.38	513
WBS000271	5.20	423
WBS000167	2.84	430
WBS000128	1.89	394
WBS000091	1.18	372

Table A5. Top ten teams handling the most Incident activities; enrichment factor is defined as the number of activities in the top 1% o Incidents divided by the number of activities across all Incidents.

Team	Relative frequency for all Incidents (%)	Team	Relative frequency for top 1% of Incidents (%)	Enrichment factor
TEAM0008	18.4	TEAM0003	12.86	26.5
TEAM0039	4.0	TEAM0002	11.48	39.0
TEAM0031	3.95	TEAM0015	8.41	10.9
TEAM0018	3.77	TEAM0171	8.38	28.7
TEAM0023	3.60	TEAM0017	4.36	9.5
TEAM0007	3.36	TEAM0008	4.19	0.7
TEAM0086	2.91	TEAM0007	3.98	3.6
TEAM0075	2.65	TEAM0018	3.45	2.8
TEAM0191	2.54	TEAM0013	2.75	8.5
TEAM0015	2.36	TEAM0069	2.51	4.7

Table A6. Top 10 teams and Service Components for top 1% Incidents matrix (percentages by total Incidents per team).

	Service Component										
	162	**088**	**073**	**072**	**055**	**263**	**271**	**167**	**128**	**091**	**Cum.**
TEAM0003	35%	35%	3%	-	5%	-	-	-	1%	-	148
TEAM0002	23%	40%	11%	-	9%	-	-	-	-	-	129
TEAM0015	-	-	-	33%	-	-	33%	-	-	-	3
TEAM0171	36%	30%	1%	-	6%	-	-	-	-	-	132
TEAM0017	-	-	-	33%	-	67%	-	-	-	-	43
TEAM0008	-	-	10%	-	-	-	-	-	1%	20%	1,568
TEAM0007	-	-	82%	-	-	-	-	-	-	-	859
TEAM0018	-	-	-	62%	-	-	33%	-	-	-	1,065
TEAM0013	4%	3%	19%	-	-	-	-	-	-	-	72
TEAM0069	-	-	1%	-	-	-	-	12%	-	-	259

Table A7. Activities of top 1% Incidents.

Activity	Relative frequency for total Incidents (%)	Proportion of activities in top 1% Incident activities (%)
Assignment	24.47	3.99
Reassignment	18.61	5.28
Operator Update	17.3	4.49
Update	10.6	4.3
Status Change	6.72	1.86
Closed	4.3	1.18
Update from customer	3.94	19.66
Open	3.48	1.02
Caused By CI	2.59	1.03
Communication with customer	1.33	3.16
Communication with vendor	1.33	11.21
External Vendor Assignment	1.2	3.86
Description Update	0.98	3.08
Analysis/Research	0.77	13.74
Reopen	0.48	2.91
Quality Indicator Fixed	0.31	0.52
Impact Change	0.24	2.72
Urgency Change	0.23	2.57
Pending vendor	0.22	0.71
Vendor Reference Change	0.21	25.24
Resolved	0.19	1.61
Vendor Reference	0.09	1.30
Notify By Change	0.06	2.93
Quality Indicator	0.05	0.26
Mail to Customer	0.05	0.18
Quality Indicator Set	0.04	0.27
Problem Closure	0.04	15.15
Affected CI Change	0.03	2.68
Service Change	0.03	3.15
Incident reproduction	0.02	20
External update	0.02	0.21
Callback Request	0.02	4.08
Problem Workaround	0.02	1.77
Referred	0.02	8.33
Contact Change	0.01	3.45