Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

MODELING EMERGENCY CARE IN HOSPITALS: A PARADOX - THE PATIENT SHOULD NOT DRIVE THE PROCESS

Andrew M. Hay

Hospital Navigator Ltd
London, UNITED KINGDOM

Edwin C. Valentin
Rienk A. Bijlsma

Systems Navigator BV
Delft, THE NETHERLANDS

ABSTRACT

The objective in the creation of domain specific discrete event simulation environments is to facilitate model development in the chosen domain. In the creation of such environments, there has been a tendency to adopt a factory based world view. In this paper, we describe an approach to the creation of a generic modeling environment in the healthcare domain that breaks away from the conventional entity driven request for resource. Our approach has enabled us to create models of emergency care in four UK NHS hospitals that reflect more realistically the way emergency care is actually delivered. It appears, paradoxically, that in simulating emergency care, it is best if the patient does not come first.

1 INTRODUCTION

The search worldwide for increased efficiency and cost effectiveness in healthcare delivery has made the healthcare environment attractive to the systems modeler. The variability and complexity of behaviours inherent within healthcare systems demand the analytic power of discrete event simulation. The potential for widespread application of discrete event simulation technologies in healthcare is a stimulus to the development of healthcare specific modeling environments aimed at enhancing modeler productivity, an example being the Medmodel environment (Harrell and Lange, 2001). In the Medmodel environment and in other recent simulation solutions proposed for emergency care (Rossetti et al., 1999; Centeno et al., 2003; Samaha et al., 2003; Takakuwa and Shiozaki, 2004; Sinreich and Marmor, 2004), it has apparently been most convenient to employ a world view which sees the hospital as a factory variant through which the patient passes, claiming whatever resources are required to perform the process of the moment. The patient is the driver. We have discovered, however, that this way of thinking can produce perverse results. It is our purpose, therefore, to propose an alternative world view in which the medical resource rather than the patient entity is the driver.

2 WHY THE PATIENT AS DRIVER IS AN INACCURATE REPRESENTATION

In the conventional modeling approach, entities join a queue, each with a certain priority. Often, in manufacturing systems, priority is expressed as a due date. The queue is processed according to priority. If the available resource is insufficient, some entities will be processed after the due date. This approach will not work for emergency care in a hospital, for two reasons:

1) It is not acceptable for patients with low priority conditions to wait interminably for treatment. With the conventional modeling approach, if the hospital is busy, patients with high clinical priority will be treated first and the low priority patients will be treated either not at all or not within an acceptable time frame.

2) The medical staff has a skills hierarchy whereby senior personnel, whilst capable of performing any clinical task, generally do not do so, except at very busy times, in order to be free for consultation and other duties that may actually lie outside the model. Junior personnel, on the other hand, can perform only subsets of the possible list of clinical tasks. With the conventional modeling approach, the wrong resource may do the work – a senior doctor performs a simple, relatively low skilled task whilst a trainee remains idle.

Thus, accurate simulation of the emergency care pathway throws up two specialised requirements: 1) a waiting time dependent approach to clinical priority and 2) the ability to enable the medical resource to decide what is the most appropriate task (patient) that it should deal with next.

3 A NEW APPROACH TO MODELLING EMERGENCY MEDICAL PROCESSES

There are three elements that are core to the new approach:

1. Care pathways
2. Operating priority
3. Skill sets

3.1 Care paths

Models are assemblies of individual care pathways. A care pathway is a discrete subset of the model that contains at least one process and/or decision module. The important property of the care pathway is that it provides a convenient way to allocate a specific baseline value for clinical priority to a particular section of the model. Care pathways also form distinct subunits for the collection of simulation output statistics.

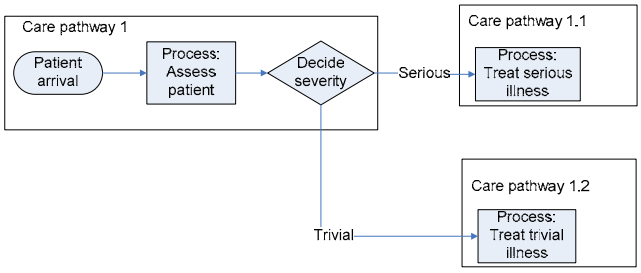

Figure 1: Example of a simple care pathway structure.

In the simple example illustrated in Figure 1, the process in the initial care pathway 1 might have a relatively high underlying clinical priority that persists into care pathway 1.1 but is reduced in care pathway 1.2. It is important to note that our approach does allow conditional override of the default priority of a care pathway in regard to individual processes within that care pathway.

3.2 Operating priority

The operating priority is a single number that expresses the value of two independent variables: 1) the underlying clinical priority of a process and 2) the length of time that the patient has waited for execution of that process. The operating priority determines the priority of a task relative to all other outstanding tasks. The operating priority is a composite score that is calculated as follows. Initially, at the moment the patient enters the care pathway, the operating priority is equal to the clinical priority of the care pathway. However, as soon as the patient joins a queue the operating priority gradually increases whilst the patient waits. Its value increments by 1 each time a specified period of waiting is exceeded. Thus, a patient awaiting resuscitation might have an initial priority of 5 because the care pathway for resuscitation has a clinical priority of 5. The user might have specified that the operating priority should increase by 1 every 2 minutes. Thus, after a 10 minute wait, the value for the operating priority would have climbed to 10. By contrast, a patient awaiting treatment of a simple sprain might have an initial operating priority of 2 and the operating priority might only increment by 1 every 15 minutes. This patient would have to wait 2 hours before the operating priority would reach a value of 10.

3.3 Skill sets

Each process requires the application of a particular skill. Instead of making direct claim to a particular kind of resource, patients claim a skill relevant to the process to be performed. The skill is to be found within the skills lists that characterise each individual type of resource. The model searches each skills list in order. The further down a list a skill is positioned, the less likely it is that the resource owning that skill will be claimed.

A more important way of protecting individual resources against inappropriate claims is provided by the threshold value for operating priority that qualifies each skill in a skills list. The value for the operating priority of the outstanding task must equal or exceed the threshold value before the resource owning the skill can be claimed.

The two skill sets illustrated in Table 1 show how this arrangement works. The senior doctor is able to perform all the tasks that the junior doctor can perform, but it is better if he or she is not generally called to perform certain tasks - in this example, the last three tasks in the list - that are also within the capability of the junior doctor. Thus the threshold value for the operating priority for these three tasks has been set high, at 10, 10 and 15 respectively. By taking account of the clinical priority of the care pathway and the rate at which the operating priority increments with waiting, it is possible to control the length of time for which the patient is to be delayed before a senior doctor is deployed for treatment.

Senior doctor	
Skill	***Threshold***
Treat_Resuscitation	1
Treat_RapidAssessment	1
Treat_ConsultWithJunior	1
Treat_StandardAssessment	10
Treat_ContinuingCare	10
Treat_SutureLaceration	15

Junior doctor	
Skill	***Threshold***
Treat_Resuscitation	1
Treat_StandardAssessment	1
Treat_ContinuingCare	1
Treat_SutureLaceration	1

Table 1. Examples of skill sets for doctors of different seniority.

3.4 Summary

Figure 3 summarises the key elements of our approach. The sequence is as follows. The patient enters the queue for the process "Assess patient". The requirement for the appropriate skill, in this case "Treat_StandardAssessment", is added to a single queue – the so-called "ActiveResourceQueue" – that exists for all outstanding requests that there are currently for active[1] resources throughout the model. The request is added to the middle of the ActiveResourceQueue according to the initial value of the operating priority, which is the clinical priority of the care pathway and is 3 (Step 1). The patient waits for 20 minutes without being allocated the appropriate resource. During this time, the operating priority has climbed to 5 and the model promotes the request to the top of the queue, as shown in the second, lower ActiveResourceQueue picture (Step 2). The model now searches for any resource with the necessary skill and for whom the threshold value for the operating priority has been equaled or exceeded (Step3) and finds a suitable doctor.

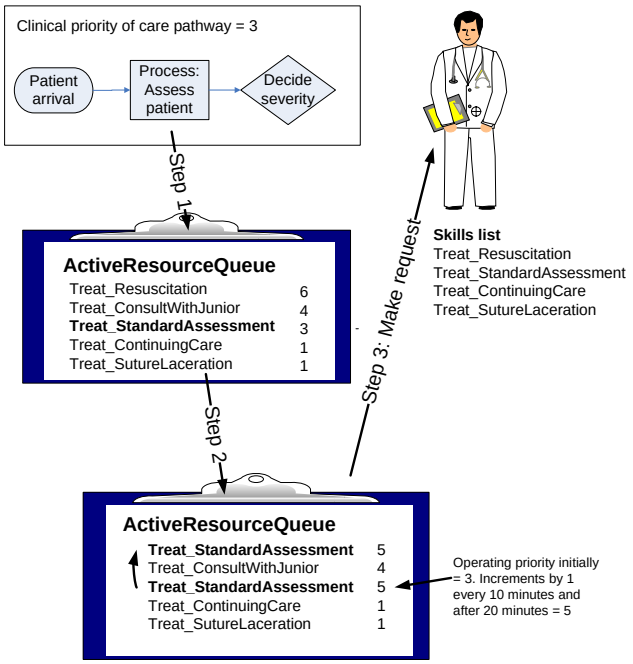

Figure 3. The logical sequence that is followed in matching a task to a medical resource. See text for explanation.

It is by using skill sets and operating priority thresholds, supported by the device of a wait dependent operating priority score, to drive resource allocation that we have been able to simulate the kind of balanced, graded response that healthcare professionals adopt in real life. There is generally a spectrum of resource that can perform a particular task. Exactly which resource is allocated depends on how severe the patient's condition is, how busy the hospital is and for how long patients have been waiting for treatment.

4 IMPLEMENTATION

4.1 Arena template "Medical Process"

These concepts have been implemented in a set of Arena building blocks. The building blocks simplify the construction of models for the emergency care pathway in hospitals. There are 10 building blocks and they have been assembled within an Arena template, called "Medical Process". Each building block has its own user interface for easy configuration.

4.2 Excel interface for model configuration and output

We have found that models of emergency care processes can be complex. Because of this we have created an Excel interface that enables configuration of the simulation model from a single location and also collects all the output data from the simulation model.

5 RESULTS

In this section we report the results of two different simulation experiments. The first set of experiments compares the outputs from the same simple model assembled 1) according to conventional principles in Arena and 2) using our new concepts and building blocks. The second set compares the outputs from a realistic model of an A&E department assembled using the new concepts and building blocks, under two conditions: 1) with our new devices operating fully and 2) with the new devices rendered inoperative.

5.1 Comparison of outputs from the same model built "conventionally" and with the new building blocks

The model contains just two processes, A and B, which run in parallel. Process A receives 70 patients each day. These are low priority patients and can be dealt with by either of two junior doctors or by a senior doctor. The process time varies between 30 and 90 minutes. Process B receives 10 high priority patients each day. These can only be dealt with by the senior doctor, who requires between 10 and 30 minutes to complete each transaction. Table 2 compares the outputs from three different simulation runs.

[1] We distinguish between active and passive resources. Active resources are those involving human intervention. Passive resources are provided by physical equipment such as beds. Both resources are handled in the same way. For the sake of simplicity we deal only with active resources in this paper.

	Conventional Arena model	**Operating priority device off**	**Operating priority device on**
	Minimum - Average - Maximum		
Utilization of junior doctors	100% - 100% - 100%	100% - 100% - 100%	91% - 94% - 98%
Utilization of senior doctor	11% - 13% - 16%	12% - 14% - 16%	94% - 96% - 100%
Wait time for process A (hours)	12 – 15 – 18	13 – 14 - 16	1.1 – 2.6 – 5.1
Wait time for process B (hours)	0 – 0.03 – 0.15	0 – 0.03 – 0.13	0.4 – 0.5 – 0.8

Table 2. Comparison of outputs from 3 different formulations of same model (see text). The results are tabulated in triads to represent, respectively, the average of the minimum values, average values and maximum values seen in 20 replications.

The compared run outputs are:

1. Model built using conventional Arena modules. In this formulation, the senior doctor performs process A if he is available to do so, but responds preferentially to the higher priority requests from process B.
2. Model built using Medical Process modules, with operating priority device switched off.
3. Model built using Medical Process modules, with operating priority device switched on. In this formulation of the model, the senior doctor attends to patients in the following order of priority:
   (a) Patients who have been waiting for process B for more than 20 minutes.
   (b) Patients who have been waiting for process A for more than 1 hour.
   (c) Any patients waiting for process B

When the operating priority device is switched off, the model built with the Medical Process modules behaves almost identically to the model built with conventional Arena modules. When the operating priority device is switched on, waiting times for process A are dramatically reduced at the cost of a relatively small increase in waiting times for process B, and there is more effective and appropriate utilisation of the senior doctor.

5.2 Comparison of outputs from a model of an Accident & Emergency (A&E) department operating with and without the new principles

In this section we show how application of the new modeling environment to simulation studies carried out in four separate NHS A&E departments (NHS A&E departments in the UK are broadly equivalent to the Emergency Rooms [ER] found in the USA) has enabled realistic representation of working behaviours that would not otherwise have been possible.

Apart from the obvious requirement that all patients should be treated expeditiously and to a high clinical standard, A&E departments in the UK have now to meet a national standard to the effect that no patient should remain within the A&E department for more than 4 hours, being discharged or admitted during that time.

The particular difficulties of the A&E department lie in the variability of patient arrival rates and in the content of the caseload. Figure 4 shows data for the arrivals at an A&E department during a single month and the variation in arrival rates from day to day. Furthermore, the workload conferred by each arrival also varies, from trivial to substantial, from minor cuts and sprains to life threatening conditions requiring immediate and intensive resuscitation. This variation in individual case content has the effect that there is only a weak correlation between the number of arrivals on any one day and the effort required to deal with those arrivals. Figure 5 shows how the numbers of patients breaching the 4 hour target related to arrival rates at one

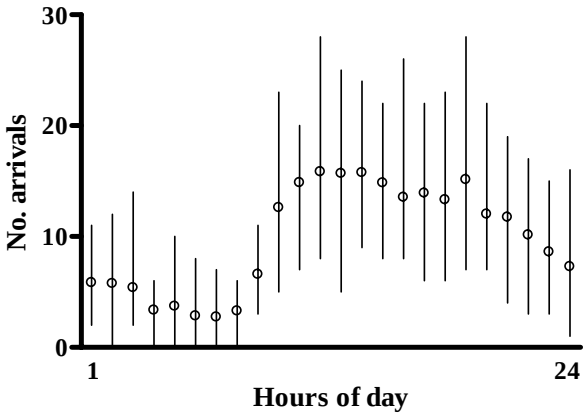

Figure 4. Patient arrivals during each hour of the day at an A&E department (Hospital A) during January, showing means and ranges.

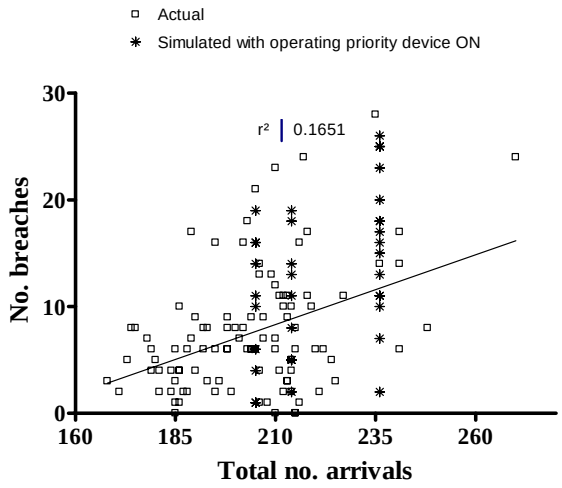

Figure 5. Plot of the numbers of patients breaching the 4 hour performance target against the numbers of arrivals during 24 hours. Observations made in the A&E department at hospital A during a 4 month period and results from a simulation model of the A&E department at hospital A.. The linear regression line is for the "Actual" data and has a slope that is significantly not zero.

A&E department during a 4 month period ("Actual" values). Whilst the regression line has poor predictive powers (regression coefficient = 0.1651), the slope of the regression line is significantly not zero, consistent with the loose correlation between arrival numbers and real workload that is to be expected.

A&E medical teams have become adept at managing the conflicting pressures exerted by those patients who are seriously ill and those who have minor complaints but are close to the 4 hour target. Nevertheless, many departments are criticised by the standards authorities because they fail intermittently to meet the requirement that 98% of patients are dealt with fully within 4 hours. During the past 2 years, we have worked with four such departments who have turned to simulation to find a remedy for such failure.

Obviously, in simulating the workings of an A&E department, many factors must be considered. Purely within the department itself, are medical and nursing staff optimally disposed throughout the 24 hours and is there efficient use of space? And, in regard to the department's dependency on external factors, how readily available are second opinions from specialist staff and beds for patients requiring admission? All these factors have been considered in our simulation models, but in this paper we wish to focus specifically on the issue of how best to simulate the way in which medical staff balance their efforts in dealing with competing clinical demands. In the A&E departments with which we have worked, we have generally demonstrated an under provision of medical staff as a cause of breach of the 4 hour target. If an A&E department is to make a successful case for the employment of more medical staff, any model used in the argument must provide a completely convincing representation of the way in which staff is utilised.

Figure 6 shows the results of a series of simulation experiments in which we used, as input to our simulation model, the downloaded file for the record of arrivals at hospital A on 3 separate days, chosen to represent the 50th (204 arrivals), 75th (214 arrivals) and 95th centiles (236 arrivals) of workload as judged (with the limitations referred to above) by the arrival rates that were observed during a four month period. The model was used in two configurations: 1) with the device for managing operating priority in play and 2) with the device switched off. (In the "off" state, the model continues to use skill sets to allocate resource, but all skill thresholds and the wait induced increment in operating priority are inoperative.) The regression line for the breach rates obtained with the operating priority device in play is close to the regression line calculated for data actually observed. On average, the model seems to predict 2 or 3 more breaches than were actually observed, an insignificant number. The regression line for breach rates obtained with the device switched off is distinctly separate, predicting an average 12 to 20 more breaches than were observed. This is a large and significant discrepancy that indicates that the model in the "off" state does not deploy medical staff effectively.

Using the Mann-Whitney test in a comparison of the breach rates obtained from the two configurations of the model provides P values of 0.0827, 0.0144 and <0.0001 for

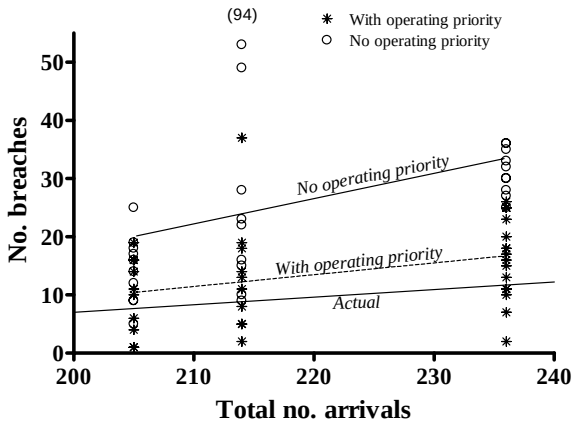

Figure 6. The effect of using the operating priority device on computed breach rates. Each point represents the number of patients exceeding 4 hours in the A&E department model. The regression line labeled "Actual" is the line obtained from analysis of the data in Figure 2 – the individual data points have been removed for the sake of clarity. The other two regression lines are derived from the data points in the figure. The item (94) refers to a data point that is off the scale of the Y axis and has a value of 94.

the workloads related to 204, 214 and 236 arrivals respectively. It seems that as the medical staff in the model become busier, the more apparent is the effect of the operating priority device.

Figure 5 shows how the simulated breaches correspond to actual breaches when the operating priority device is switched on.

It is rare for breaching patients to remain within the A&E department for more than 5 or 6 hours in total. Figure 7 shows that when conventional methods of simulating priority are used (the operational priority device is off), patients can remain in the system for nearly 24 hours – the length of the simulation run and a circumstance that is never observed in real life. The difference is probably greater than that shown in Figure 7 because, on average, 10 fewer patients of the 236 entering the model passed completely through the system when the operating priority device was off.

6 DISCUSSION AND CONCLUSIONS

The business requirement in regard to A&E departments in UK hospitals is that such departments should operate consistently within the 4 hour performance target. This is difficult for many departments when confronted by the higher levels of a widely variable clinical demand. From the simulation point of view, it is important to provide convincing representations of department behaviour at these high levels of demand in order to develop credible strategies for improving performance. Realistic representation, however, of A&E department behaviours provides a particular challenge for the simulationist. This is because the medical staff varies its working methods depending on the degree to which the department is busy. The medical staff knows how to make best use of the differing skills of junior and senior doctors, balancing the use of differing levels of clinical expertise against the complex needs of patients with a wide range of clinical priorities.

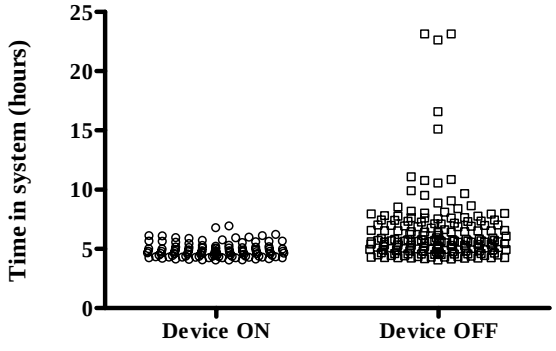

Figure 7. The effect of the operating priority device on length of stay in the A&E department. Cumulative data obtained from 5 simulations in each configuration, using as model input 236 arrivals in 24 hours. The difference between the two datasets is statistically highly significant ($P<0.0001$, Mann-Whitney test).

If accurate guidance in regard to the optimal deployment of clinical staff is a key objective for simulation work in A&E departments, the representation of the way in which the medical staff directs its efforts must be realistic. In this paper we have shown that simulation of the way in which clinical resource is allocated within an A&E department demands a more complicated approach than that adopted in the conventional factory world view. In this more complicated approach, in which resource allocation is mediated through the use of skill sets and operating priority, the patient entity is no longer the primary driver. Instead, the allocation of resource is guided by complex drivers that encompass ideas of relative suitability for the task in hand and even handedness towards a widely varying patient demand.

We have implemented these concepts through simulation models that have been built using Arena and by means of a template that contains building blocks with the necessary logic. Models of four separate UK A&E departments have provided proof of concept in that we have been able to replicate failure rates against the 4 hour performance target with a precision that would have eluded us had we used a conventional approach to the claiming of resource.

Finally, the work presented here carries clear implications for projects involving the development of domain specific tools to aid productivity in the construction of healthcare simulation models.

REFERENCES

Centeno, M.A., R. Giachetti, R. Linn and A.M. Ismail. „A simulation-ILP based tool for scheduling ER staff.“ *Proceedings of the 2003 Winter Simulation Conference*, pp.1930-1938, 2003.

Harrell, C.R.; V. Lange. “Healthcare simulation modelling and optimisation using Medmodel” *Proceedings of the 2001 Winter Simulation Conference*, pp.233-238, 2001

Rossetti, M.D, G.F. Trzcinski and S.A. Syverod. “Emergency department simulation and determination of optimal attending physician schedules. *Proceedings of the 1999 Winter Simulation Conference*, pp.1532-1540, 1999.

Samaha, S., W.S. Armel and D.W. Starks. “The use of simulation to reduce the length of stay in an emergency department.” *Proceedings of the 2003 Winter Simulation Conference*, pp.1907-1911, 2003.

Sinreich, D. and Y.N. Marmor. “A simple and intuitive simulation tool for analyzing emergency department operations.” *Proceedings of the 2004 Winter Simulation Conference*, pp.1994-2002, 2004.

Takakuwa, S. and H. Shiozaki. “Functional analysis for operating emergency department of a general hospital.” *Proceedings of the 2004 Winter Simulation Conference*, pp.2003-2011, 2004.

AUTHOR BIOGRAPHIES

ANDREW M. HAY, for 25 years a consultant surgeon in the UK NHS, has a long-standing interest in the scheduling problems of hospitals. His current efforts are focused on the development of operational decision support systems for hospitals. His email address is: <andrew.hay@dsl.pipex.com>

EDWIN C. VALENTIN is a research fellow at the Technical University Delft and simulation consultant at Systems Navigator BV in The Netherlands. His main interest is in the creation of domain specific development environments for discrete event simulation. Edwin has implemented such environments at Nestle, Sandd, and the British Home Office. His email address is: <edwin.valentin@systemsnavigator.com>.

RIENK A. BIJLSMA has 10 years' experience in Operational Research. As Managing Director of Systems Navigator BV, he has introduced simulation technology to a wide variety of industries in Europe, including ABN AMRO Bank, Nestlé, Philip Morris and Unisys Consulting (Antwerp Port Authorities).