Resilient Business Process Management: Framework and Services

Pedro Antunes 1, Hermâni Mourão 2

1 Department of Informatics, Faculty of Sciences of the University of Lisbon, Campo Grande, 1749-016 Lisboa, Portugal
paa@di.fc.ul.pt

2 Business School of the Polytechnic Institute of Setúbal, IPS Campus, Estefanilha, 2914-503 Setúbal, Portugal
hrmourao@gmail.com

Corresponding Author

Pedro Antunes
Department of Informatics, Faculty of Sciences of the University of Lisbon, Campo Grande, 1749-016 Lisboa, Portugal
paa@di.fc.ul.pt
Phone +351-21-750 0605
Fax: +351-21-750 0084

Abstract

This research contributes to extend Business Process Management (BPM) systems with resilience support. We applied concepts derived from resilience engineering and the study of hazards in highly reliable organizations to characterize resilient BPM. We developed a resilience framework based on two criteria, control, which may be prescriptive, mixed or discretionary, and response, considering planned and non-planned actions. We review and classify techniques developed in the BPM field dealing with various types of hazards. Three out of five categories involve humans in various ways. A special focus is given to discretionary/unplanned human interventions. We developed a set of services integrating resilience support in BPM systems, including detection, diagnosis, recovery and escalation. One important feature associated with the diagnosis service is handling the dynamic trajectory of hazards. Another fundamental feature provided by the escalation service is involving different operators in the collaborative activities necessary to overcome more complex hazards.

Keywords


1. Introduction

Various organizations optimize their business through process orientation. Business Process Management (BPM) integrates a collection of technologies capable to translate business process models into computer-supported activities, relinquishing routine management and control tasks from the organizational agents. Another goal often attributed to BPM is lessening organizational change through better isolation of functions like work coordination, resource management, communication and service decomposition.

This process orientation has however one fundamental problem: requiring organizations to formalize their business processes down to the task-level details required by BPM technology. But that rationalistic/mechanical approach is often infeasible or harmful to organizational behavior. Firstly, there is a trade-off between responsiveness and formalization. High formalization makes organizations less responsive to turbulent environments. Low formalization naturally increases responsiveness, but challenges the capacity of BPM systems to effectively coordinate business activities.

Secondly, we also find a trade-off between detail and ambiguity. Most service-oriented organizations deal with great levels of informality, variability and ambiguity (Saastamoinen, 1995). Therefore many work processes must be kept at very generic and often vague levels of detail. On the contrary, BPM systems often require detailed specifications about what, how, when, who and where activities should be executed.
Besides these relatively confined issues we should also take a broader view of the organizational forces shaping BPM technology. Several researchers observe that computerization has been increasing and organizations are becoming more dependent on computing technology (Hollnagel & Woods, 2005). All along with this increasing dependency we find out that organizations and computing technology have become more complex, adopting new transformation processes, higher temporal demands, wider distribution and span of control, increased skills levels and more intensive decision-making abilities (Hatch, 2006). The consequence of this trend is that organizations have become more prone to hazards (Perrow, 1994). Interestingly, it seems that technology itself is becoming less prone to failure, while human and organizational factors have been increasingly blamed for accidents (Hollnagel & Woods, 2005). And since people, organizations and technology converge in BPM systems, they are necessarily at the centre of the problem, not only contributing to cause accidents but also offering opportunities to tackle accidents (Sell & Braun, 2009).

Resilience engineering is the research field aiming to understand the complexity associated with socio-technical systems while studying methods, techniques and tools to increase the organizations’ capacity to maintain operations when facing accidents (Hollnagel, Woods, & Levenson, 2006). In this paper we apply the resilience concept to BPM systems. We develop a resilience framework based on two criteria, control and response, adapted to BPM. We also review the techniques developed in the BPM field analyzing how they cope with various types of hazards, ranging from component failures to large-scale hazards. Our review is structured to highlight the BPM support to increasing resilience levels.

In this paper we also tackle the services and information models necessary to build resilient BPM. One aspect that has received significant attention is the support to collaborative activities necessary to handle large-scale hazards. Flexibility, decision-making and collaboration are intrinsic characteristics of resilience and therefore resilient BPM has to integrate collaboration support.

The paper is organized in the following way. In section 2 we discuss the fundamental requirements of resilient BPM. In Section 3 we propose a resilience framework. In section 4 we apply the framework to review the existing resilience support in BPM systems. Section 5 describes a collection of core services implementing resilient BPM. Finally, in sections 6 and 7 we discuss the approach and present the conclusions from this research.

2. The Fundamental Requirements of Resilient BPM

According to Perrow (1999), the interactive complexity and tight coupling between people and technological components of organizational systems has been increasing, which leads to unpredictability of operations and inevitably to accidents. In that sense, accidents should be considered “normal” in complex systems operations, a theory that has become known as NAT (Normal Accident Theory (Perrow, 1994)).

Accidents arise from the combination of hazards with holes in defenses caused by active failures (unsafe acts) and latent conditions (Reason, 2008). In order to deal with accidents, the system developers should work out various defenses, barriers and safeguards (Cacciabue, 2004).

Resilience is a property intimately associated with the organizations’ capacity to avoid, contain and mitigate accidents. A deep understanding of resilience is emerging from the study of many High-Reliability Organizations (HRO) such as nuclear power production, aviation, space exploration, healthcare, air traffic control and chemical production (Gauthier, Davis, & Schoenbaum, 2006; Perrow, 1999). The major interest in HRO comes from their capacity to achieve high performance while operating in hazardous conditions (Weick & Sutcliffe, 2001). Of course achieving high performance in these conditions leads to some distinctive behaviors. A fundamental one concerns safety: HRO may operate beyond the envelope while avoiding human injury and preserving environmental and financial assets. Another fundamental behavior is sensitivity to operations (Weick & Sutcliffe, 2001): HRO are capable to deal with unexpected events, latent failures, losses in defensive barriers and, ultimately, catastrophic accidents. And we shall also refer another fundamental behavior, combining humans and technology, which is making decisions under fluid and complex circumstances, and lack of time, resources, knowledge and experience.

Therefore resilience is not a technological or organizational property but a combination of both. It is the combination of technological features, such as redundancy, protection systems and good engineering design (Leveson, Dulac, & Marais, 2009) with organizational features such as sensemaking (Weick, 2001), training,
and decentralized decision making, that builds up what is commonly designated resilience. The recent studies on resilience clearly emphasize the integration between the organizational and technological views in complex socio-technical systems (Hollnagel, et al., 2006):

- Supporting various levels of severity, ranging from simple failures of key resources to catastrophic accidents (Turoff, Chumer, Van de Walle, & Yao, 2004);
- Supporting the coexistence of stable processes with unstable changes in the operating environment;
- Supporting the dynamic construction and update of situation awareness, i.e. perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and projection of their status in the near future (Endsley, 1995);
- Supporting knowledge representation and management, a fundamental drive to decision making;
- Supporting flexible operations and unplanned tasks whenever necessary, while deferring authority to the most adequate persons, often the ones operating at the sharp end (Leveson, et al., 2009);
- Supporting the opportunity to experiment with and learn from the novel, innovative and challenging situations that emerge from hazards; and
- Supporting the transition from emergent to normal operations.

It may be argued that BPM does not aim to support emergent situations. Indeed, the origins of BPM came from the practical objective to leverage existing technical assets in organizations, in particular document automation and distributed systems (Borghoff & Schlichter, 2000; Nutt, 1996). The original literature on BPM is dominated by a hard-systems perspective where the control of the organization resides in the machine (Melão & Pidd, 2000). According to that perspective, BPM manages organizational activities based on predefined procedures.

Further analyzing the hard-systems perspective, a BPM system coordinates a collection of human and automated activities tied together by a set of precedence relations and pursuing a common organizational goal (Sheth, et al., 1996). The coordination of activities is most often based on a specification (also designated process model (WfMC, 1999)) or, in some cases, by ad-hoc decisions made by humans participating in the process (typically designated ad-hoc workflow (Georgakopoulos, Hornick, & Sheth, 1995)). These definitions represent a simplification of the coordination mechanisms adopted by many service oriented enterprise architectures (Arora & Nirpase, 2008; Leymann, 2002), but such simplification is accomplished with the purpose to highlight the aspects of coordination and control that are central to our discussion.

The infrastructure necessary to deploy BPM systems includes two levels of human intervention: (1) as system developers, responsible for analyzing the organization and synthesizing a collection of process models later on instantiated by the BPM system; and (2) as process participants, with the responsibility to accomplish the designated activities according to precedence relationships specified by the developers and managed by the system. But these two levels of human intervention are significantly constrained. Starting with the system developers, they are constrained in the following ways:

- System developers may not fully understand the tightly coupling and complex interactions between the various technological components existing in the organization and the BPM system; and regarding the BPM system per se, they may also not fully understand the relationships between the human and automated activities necessary to accomplish the designated organizational goals;
- Even when some of these intricacies are understood, they may not end up being specified in the process models, most often to avoid turning the process models incomprehensible or unmanageable;
- System developers may naturally specify erroneous process models (Casati, Ceri, Paraboschi, & Pozzi, 1999; Heindl, 1998);
- And we finally have also to consider that system developers are often constrained by time. The analysis of organizational processes, specification of process models and subsequent instantiation takes time, which may not be available when the environment is turbulent and the organization behavior must adapt at a fast pace to such changes.

Regarding the process participants, we may also find some important constraints to their roles in BPM systems:

- The main constraint is related with what is designated model consistency. Most BPM systems, if not all, require that process activities be executed from the start to the end point without leaving any dead-locks or live-locks or activities never being triggered (van der Aalst, 2001). Model consistency
implies that any ad-hoc changes, in particular the ones carried out by humans, should be prevented from putting the BPM system in an inconsistent state (Faustmann, 2000; Jorgensen, 2001; van der Aalst & Basten, 2002). This clearly means that if model consistency is mandatory then there is some degree of control from the technology over what human activities are allowed or not allowed to be carried out;

• And another constraint is related with the impact of automation on human behavior. In the one hand, we have to consider that a significant portion of the coordination and decision-making abilities have been transferred from the humans to the technology, in the form of process models. And in the other hand, we also have to contemplate that, by transforming work into a collection of black-box activities, humans may loose perspective of the whole operations carried at the group and organizational levels. This situation has some similarities with what has been designated in aviation as the out-of-loop problem and glass-cockpit syndrome (Marianne, 2000; Redmill & Rajan, 1997), and more generally as automation surprises (Woods & Hollnagel, 2006).

Many commercial BPM systems failed to address the above constraints, which has led to a difficult acceptance by their hosting organizations (van der Aalst & Berens, 2001; B. Weber, M. Reichert, & S. Rinderle, 2008).

The importance of the human role in BPM has however been increasingly recognized and is carrying BPM out of the hard-systems perspective towards a more eclectic view integrating humans and technology (A. Agostini & G. De Michelis, 2000; Brahe & Schmidt, 2007).

Suchman (1987, 2005) ignited this trend. She investigated office automation from a distinctive standpoint, sociology, analyzing in particular the inference, interpretation and contextualization often necessary to carry out process activities. In quite a strong statement, Suchman (1993) questioned the conventional order, compliance, focus on efficiency and technology-driven agenda imposed by office automation. The major argument was that control should reside in humans and not in the technology. Others defended the BPM origins as having fewer prejudices than avowed by Suchman. For instance, the hard-systems support contributes to document what occurs in complex information systems, making work structures visible, and also improve coordination and accountability (Winograd, 2006).

The major outcome from this debate is a modern view of BPM, bearing in mind the complementary roles of automation and discretionary human behavior, the former offering guidance and accountability, and the latter contributing with openness and flexibility (Bannon & Bødker, 1997; Grinter, 2000; Herrmann, Hoffmann, Loser, & Moysich, 2000; Herrmann & Loser, 1999; Taylor & Virgili, 2008; van der Aalst, 2005). This view opens the opportunity to synthesize the major properties of resilient BPM.

Bringing again the major assumptions of NAT, accidents are inevitable in complex systems, typically caused by small errors interacting in unexpected ways and cascading in increasingly larger failures, which may end up with overall system failure (Perrow, 1994). Thus it is very important that BPM systems maintain business operations under the occurrence of small errors and possible cascading events. The ability to adjust the BPM system to the actual operational conditions, applying preventive, containment and mitigation measures to different hazardous situations is therefore a core BPM property.

Researchers have recognized some aspects of this problem since the early days of office automation. For instance, the need to develop high-level languages conciliating process modeling with change management was early identified (C. Ellis & Nutt, 1980; Hammer, Howe, Kruskal, & Wladawsky, 1977). The need to support enterprise-wide, heterogeneous, autonomous and distributed operations was also identified (Bussler, 1999; Worah & Sheth, 1997). And researchers have also developed various techniques to improve robustness and flexibility (van der Aalst, Basten, Verbeek, Verkoulen, & Voorhoeve, 1999; van der Aalst & Berens, 2001).

Resilient BPM requires robustness to avoid errors, e.g. through better system development and better process models; and flexibility to adjust the operations to deviations between the process models and the existing conditions. Though it should be emphasized that robustness is quite challenging for BPM because predicting every possible hazard during the development phase is considered very difficult or even impossible and makes the systems very complex and hard to manage (Casati, 1998; Dayal, Hsu, & Ladin, 1990; J. Eder & Liebhart, 1998; Klein & Dellarocas, 2000; Mohan, Alonso, Guenthoer, & Kamath, 1995). The problem with flexibility is that it requires process participants to intervene in the system beyond the mere accomplishment of their formally assigned tasks. And so flexibility requires dealing with the limitations imposed by model consistency, automation and black-boxes.
A good balance between flexibility and robustness should nevertheless be envisaged (Nomura, Hayashi, Hazama, & Gudmundson, 1998). Robustness is important to keep the organization under control. And flexibility is necessary to react to hazards. The main objective of resilient BPM systems could then be summarized as supporting flexibility without losing all the advantages of BPM automation. In Figure 1 we give a summary view of the major requirements associated with resilient BPM discussed above.

Figure 1 – Summary view of major requirements associated with resilient BPM.

3. Framing the Problem

Before we move on to assess the current status of BPM technology regarding the resilience property, we first have to articulate a classification framework. This framework should incorporate the major concerns and requirements discussed in the previous section, namely the integration of humans and technology, and robustness and flexibility support.

To build our framework, we adopted a model developed by Reason (2008) to analyze human performance. According to Reason, human performance depends on two dimensions governing human action: situation and cognitive control. The first dimension expresses the increasing complexity of the problem situation: (1) routine, the performer immediately recognizes the problem and has the skills necessary to accomplish the task; (2) trained-for, the performer has been trained to recognize and handle the problem according to known procedures and best practices; and (3) novel, when there is no prior knowledge about the problem situation.

The cognitive control concerns the mental modes necessary to handle the problem: (1) automatic, which means the task may be executed in an almost unconscious way; (2) conscious, when the mental activities must be made explicit to guide action; and (3) mixed, when the performer swings between the conscious and unconscious modes.

The combination of these two dimensions leads to three performance levels designated skill-based, rule-based and knowledge-based, the reason why this is known as the skills-rule-knowledge framework (Reason, 1990). The skills level addresses unconscious tasks accomplished by humans when facing routine work situations. We find rule-based performance in situations where tasks have been planned and prescribed to workers but giving them decision latitude concerning the details. And knowledge-based performance is found whenever workers find novel situations where their decision-making abilities must be fully exercised.

One interesting aspect of this framework, and the major reason why we adapted it to BPM systems, is that its extrapolation to organizational work is quite straightforward. We find that organizational control may range from prescriptive to discretionary. In between we find mixed control situations relying on prescriptive and discretionary actions. This classification is adequate to BPM since the prescriptive actions are fundamentally related with process models and activity coordination, while discretionary actions concern the local execution of the activities managed by the BPM system. The mixed model serves to address the situations where control must flow between the technology and the process participants.
Concerning the dimension of problem situation, we made a simplification of Reason’s model to consider two types of response: planned and unplanned. The former indicates the organization has capacity to resolve the problem in a planned way, taking time and resources to analyze the problem, find a solution and develop an action plan to activate the solution. The unplanned situation concerns the cases where the organization does not have enough time to plan the solution, for instance, because there is an emergency situation. The introduction of the unplanned element in the framework not only serves to express our preoccupation with sensitivity of operations, previously identified as characteristic of HRO, but also to express our view that flexibility requires timely responses to hazards accomplished by process participants.

Figure 2 – Adopted framework to characterize resilient BPM.

We finally observe the proposed framework allows us to equate resilience as the increasing capacity to move the operations from a planned/prescriptive mode towards an unplanned/discretionary mode. We understand that HRO have a sustainable capacity to operate in all modes, and posit that BPM systems should support that capacity. The Resilience line shown in Figure 2 represents the space of allowable operations according to control and response. The bottom-left triangle shows that less demanding responses require planned interventions accomplished with prescriptive control. The top-right polygon in Figure 2 is where it is expected the response would lie when the system’s resilience is pushed to its limits. In these circumstances, the discretionary mode of control may be necessary to cope with unstable changes in the operation environment or different levels of accidents severity. In the next section we revise the BPM literature according to this framework.

4. Review of Resilience Support in BPM

In the previous section we proposed a framework characterizing BPM resilience according to planning and response dimensions. We will now use the framework to review how BPM systems have been supporting resilience. Five types of support were identified.

1. Failure handling. BPM systems operate in heterogeneous, distributed and autonomous platforms that are prone to component, communication and system failures (Bussler, 1999; Worah & Sheth, 1997). Two different types of failures have been identified in the literature (J. Eder & Liebhart, 1995): (1) basic failures, associated with malfunctions in the underlying technological infrastructure including networking, database management and operating system; and (2) application failures, provoked by unexpected processing and data inputs/outputs;

Various approaches exist to overcome basic failures. For instance, transaction-processing techniques, developed in the database management field, guarantee data integrity when the system fails. In fact, most of the commercially available databases implement the necessary transaction-processing mechanisms to react in case of failure, returning the system to a coherent state and enabling forward execution (Casati, 1998).
Another approach is using failure tolerance techniques, based on data replication and synchronization to recover from failed services without losing data (Alonso, Hagen, Agrawal, El Abbadi, & Mohan, 2000).

Application failures are somewhat more difficult to handle because they may have direct and indirect impact on the associated business semantics. BPM systems deal with processes and activities that may span over long periods of time (long running activities (Dayal, et al., 1990)). In these cases, applying the isolation and atomicity properties of traditional transaction-processing techniques may compromise the levels of concurrency and cooperation necessary to accomplish work in organizations. Therefore, Advanced Transaction Models (ATM) using relaxed Atomicity, Consistency, Isolation and Durability (ACID) properties have been developed to overcome application failures (Chen & Dayal, 1996; Georgakopoulos, et al., 1995; Jin, Rusinkiewicz, Ness, & Sheth, 1993). For instance, by relaxing the isolation property, other tasks are able to access data before a transaction finishes. Compensation tasks may then be defined for each committed task to allow backward recovery, restoring data integrity and proceeding with forward execution.

However, experiments with ATM showed a limited ability to model modern organizational contexts. According to Alonso et al. (1996), ATM solutions are biased from a database view of organizational work, which may restrict the organizational behavior. Worah and Sheth (1997) also emphasize the need to look beyond transactional processing, since it addresses a restricted application domain. Although recognizing the ATM limits, it is important to emphasize it has a strong theoretical basis to assure data integrity, model correctness and recovery on the occurrence of failures in transactional environments. This research trend was very important during the 1990s when many important developments were proposed (Worah & Sheth, 1997).

Failure handling techniques offer the lowest-level support to resilience, focusing especially on robustness. Since they are built in the BPM system during the development time, they must be planned in advance. Considering that they function in a completely automatic way, we classify them in the prescriptive category. They are the first line of defense against hazards.

2. Exception handling. We define exceptions as hazards that were predicted by the system developers during the development cycle. Unlike failures, which result from system malfunctions, exceptions come from semantic discrepancies between the actual organizational environment and the processes modeled by the system.

Various solutions have been devised to handle exceptions. Some rely on triggers to initiate predefined exception handlers (Casati, 1998; Chiu, Li, & Karlapalem, 2001; J. Eder & Liebhart, 1995; Luo, 2001; Sadiq, 2000). Dayal et al (1990; 1991) proposed the ECA (Event-Condition-Action) rules to separate the identification of exceptions from their handling. Casati et al (1999) extended ECA with a specification language capable to identify four types of events:

- Workflow – Triggered when a process or activity fails;
- Data – Associated with data errors in a specific activity or in a set of activities;
- Temporal – Triggered when a given time stamp is reached or a time interval is not respected; and
- External – Activated by external resources, including human intervention.

In this work, the authors also developed the Chimera-Exception language to specify exceptions and handling procedures. In a more recent work, Combi et al. (2006) extended the XPDL\(^1\) standard to include exception handling capabilities using the Chimera-Exception language.

Luo et al. (2003) characterize other types of events resulting from cross-organizational hazards: contract cannot be fulfilled, may be compromised, needs to be modified and needs to be terminated. The action part of ECA may execute several primitives belonging to two categories: data modification and process management. The former stands for operations related to data creation, modification and deletion; while the later include functions such as notifying the liable persons, starting new activities and processes, or reassigning activities to different persons. ECA rules have been used in the ADOME system (Chiu, et al., 2001) to define an object-oriented exception handling procedure. A hierarchy of rule sets was defined so that an exception would progress through each set until resolved. This approach allows applying different contexts to exception

\(^1\) XPDL is an XML-based Process Definition Language issued by the Workflow Management Coalition.
handling ranging from the more specific (a particular activity) to the more generic (a type of event), thus offering additional flexibility.

Luo et al. (2003) proposed a Case Base Reasoning (CBR) approach to extend exception handling. The main concept is to maintain a case repository with information about previous exceptions and handling procedures; and, whenever an exception is detected, automatically consult the case repository to find similar cases. By using similarity reasoning, the system enlarges our notion of planned/prescriptive resilience. The issue is that being able to learn from past events leads exception handling towards the unplanned/prescriptive mode, even if actions are not totally new. In any case the exception handling is applied in an automated way. Interestingly, Luo et al. (2003) have also proposed integrating human intervention whenever the system is unable to find an adequate match to an exception. The human intervention complements the CBR approach by assuming the responsibility for retrieving a procedure from the case repository. This approach nevertheless corresponds to an attempt to expand exception handling towards the planned/mixed mode.

3. Model adaptation. Several authors recognize the limitations of automatic approaches to failure and exception handling, which in many cases require human intervention (Casati, 1998; Chiu, et al., 2001; J. Eder & Liebhart, 1995). Human involvement is typically necessary to analyze the situation, rethink the organizational performance and adapt the organizational behavior. The model adaptations typically emerge from incomplete developments, development errors and changes in the business environment (Casati, et al., 1999; Heinl, 1998).

Model adaptation requires the capability to dynamically change the processes running in the BPM system without any disruption in the operations (Adams, Hofstede, Edmond, & Van der Aalst, 2006). The major problem that has been addressed by research is to guarantee model consistency when applying these changes (C. Ellis, Keddara, & Rozenberg, 1995; Reichert, Dadam, & Bauer, 2003; Weske, 2001). Several researchers have defined a set of rules enabling consistency checks before applying model adaptations (Rinderle, Reichert, & Dadam, 2003; van der Aalst & Basten, 2002). Two consistency criteria must be taken into consideration: structural and state-related. The former concerns schema changes and assures the new process model is consistent. The state-related criterion concerns the state of process instances that will be migrated and verifies if they may reflect the new model.

Two main model adaptation techniques have been developed (Han, Sheth, & Bussler, 1998): metamodel and open-point. The metamodel technique takes into consideration the structural and dynamic constraints to model adaptations, while the open-point technique defines special points in the process models where the adaptations can be done. The metamodel technique offers higher intervention latitude, since they do not restrict the points in the process models where the interventions may be applied. However, they require model consistency checks. In the case of open-point, the consistency checks are not necessary since the restrictions are made explicit when defining the special points. The open-point approach has the disadvantage that allowed interventions are not complete enough for some situations that require structural changes (Han, et al., 1998).

Other research lines expanding the support to model adaptations explore a Worklet Service (Adams, Hofstede, Van der Aalst, & Edmond, 2007). The main idea is allowing system developers to designate specific portions of the process models for late-binding. Only at runtime, when the designated portions are invoked, the process models must be completely specified (Weber & Wild, 2004). Furthermore, this approach allows developing a repertoire of Worklets, thus giving significant flexibility to process execution leading towards the mixed/unplanned mode. In any case the model adaptations must be executed under strict system control, the reason why we still classify these techniques in the mixed/planned mode.

4. Restricted Ad-hoc changes. In many circumstances there is no need, justification or time to plan model adaptations. This type if intervention may also result from the explicit decision to not completely model the whole complexity, detail and variations of some business processes (Heinl, 1998). Therefore ad-hoc changes concern operations not predicted in the process models and carried out during the execution phase to accomplish work (A. Agostini & G. De Michelis, 2000).

Restricted ad-hoc changes may be seen as an extension of the open-point approach previously described. The main idea is that, reaching an open-point, the operators will be allowed to specify the following actions. Restricted ad-hoc changes are typically applied to a small set of process instances and have a transient impact (Adams, et al., 2006; A. Agostini & G. De Michelis, 2000; Mourão & Antunes, 2004). This includes, for
instance, delaying an activity, designating another operator to accomplish an activity, and inserting an absent activity.

A set of exception handling patterns have been proposed by Russel et al. (2006) to handle a set of identified hazard situations joined in five distinct groups: work item failure, deadline expire, resource unavailability, external trigger and constraint violation. Various recovery strategies were then defined to cope with the concrete scenarios. These strategies usually handle the specific case and do not predict model changes. Weber et al. (2008) developed an extensive list of change patterns that BPM systems should implement to support runtime flexibility. This list may be used to compare existing BPM systems and technologies. The proposed change patterns maintain instance and model consistency, and therefore restrict human interventions.

Dourish et al. (1996) proposed Freeflow as an alternative to ad-hoc changes using constraints. Freeflow is a constraint-based modeling system that, instead of adopting process models, uses constraints to characterize work coordination. This way changing the associated constraints may change work processes. Constraint management becomes an ongoing flexible activity.

But again, the available actions are constrained by model consistency, which means a mixed control policy is necessary, combining the momentary human control crucial to understand the situation and define the following actions with the system control required to preserve model consistency. We therefore classify these techniques in the unplanned/mixed level.

5. Unstructured interventions. Occasionally BPM systems are subject to large-scale hazards with impact in the whole organization. These cascading events may occur for various reasons, including accidents, emergencies and exceedances, i.e., situations leading the organization towards the edge of safe limits (Reason, 2008). One common characteristic of large-scale hazards is they push the envelope of typical organizational decision-making by requiring timely response, lateral thinking, reinvention of work practices and collaboration.

Failure and exception handling are completely out of scope under these circumstances and may indeed pose a threat, as the study of the Three Mile Island nuclear accident demonstrated: the warning systems were incapable to supply the information necessary to realize what was happening and in fact hindered the complete understanding of the occurring phenomena (Redmill & Rajan, 1997).

It should also be considered there might not be enough time to plan model adaptations to respond to cascading events. In these situations ad-hoc changes are necessary, but should not be constrained by model consistency (Rinderle, et al., 2003). If any BPM system’s restrictions are imposed to the organization, then the organization will find workarounds outside the system (Hayes, 2000).

The BPM system should therefore support unstructured interventions under complete human control. But the support should not be limited to relinquishing control to humans. Unlike restricted ad-hoc changes, which are limited in scope, the unstructured interventions may extend to many process models and running instances, a context that may be difficult to manage without technology assistance. According to Suchman (1987), since no plan is available, human reaction should be “map” guided. Thus the major challenge implementing unstructured interventions is supporting situation awareness and guidance under emergent and evolving contexts.

Few approaches have been documented in the literature addressing unstructured interventions in BPM systems. Agostini and De Michelis (2000; 2000b) developed one such approach. As the authors state, “systems supporting articulation work must on the one hand, liberate workers as much as possible from the routine articulation work they need for coordinating themselves (script); on the other, help them to become aware of the situation where they are performing and to negotiate new cooperative work arrangements whenever a breakdown occurs (maps). Finally, they need to be open to continuous change in order to support a continuous update of their maps and their scripts.”

In the system proposed by Agostini and De Michelis, the process participants may execute the actions defined by scripts (prescriptive process models) but may also initiate “multimedia conversations” (discretionary actions) with other persons when some hazard is detected. These two components are fully integrated, allowing a conversation to be started during process execution and a process enacted during a conversation.
Another approach integrates BPM with external collaboration tools (Guimarães, Antunes, & Pereira, 1997). The purpose is to normally maintain prescriptive control but passing it to collaboration tools when a hazard occurs. However, no support was considered to continue with normal operations after resolving the hazard, neither to obtain situation awareness.

An Artificial Intelligence mechanism to help determining the type of control more adequate to handle various types of hazards has been developed by Bernstein (2000). This mechanism was also conceived to invoke decision-support tools when discretionary control is necessary. Intelligent agents have been proposed by Wang et al. (2004) on an exception handling system developed to support inter-enterprise securities transactions. A diagnostic agent collects information from several monitoring agents and investigates the nature of the problem. If the problem is recognized, the information is issued to the resolution agent that takes the necessary initiatives to resolve the problem. The authors recognize that the system can only react to situations that occur frequently.

Other developed strategies aim at supporting decision-making and are only indirectly linked to BPM systems. One case uses a knowledge base to maintain information regarding past handling procedures and to facilitate linking hazards to handling procedures (Klein & Dellarocas, 2000). Another case uses data mining to extract relevant information about the hazard and support organizational decision making (Grigori, Casati, Dayal, & Shan, 2001).

In Figure 3 we present an overview of the major techniques adopted by resilient BPM. From this overview we realize that prescriptive techniques are crucial to increase the organization’s capacity to resist to the occurrence of failures and exceptions. However, when the situation moves beyond what is codified in the system, humans become the fundamental organizational component supporting resilience.

The mixed techniques incorporate humans in prevention, containment and mitigation operations, although limited to strict rules imposed by model consistency. Model adaptations increase resilience by migrating process models towards new organizational goals. The restricted ad-hoc changes further increase resilience by allowing more immediacy and less planning. And finally, the unstructured interventions provide an increased
level of resilience by giving wider latitude of action and access to collaboration support, which seems to be a commonly adopted strategy to cope with the information and decision-making demands.

Summarizing the whole scenario, we observe that organizations must integrate various techniques covering the path from fully prescriptive to fully discretionary actions. In the next section we will discuss in more detail the mechanisms necessary to implement this view.

5. Services Necessary to Support Resilient BPM

The following discussion is derived from our experience developing some techniques discussed in the previous section, with a particular focus on the implementation of restricted ad-hoc changes and unstructured interventions (Mourão, 2008; Mourão & Antunes, 2004, 2005, 2007). Rather than focusing on a specific BPM architecture, we will characterize the high-level services necessary to implement resilient BPM across multiple architectures.

The first service we consider is the Detection Service, which is responsible for detecting the occurrence of hazards, thus addressing sensitivity to operations, previously identified as a major HRO requirement. As already discussed, various types of hazards may occur, from the most trivial to the most catastrophic, but such characterization is out of the scope of the Detection Service. We instead consider two types of detection, manual and automatic. The automatic detection is triggered by the BPM system whenever it detects basic and application failures (failures in underlying service components and inputs/outputs, respectively), and exceptions in process execution (workflow, data, temporal and exceptions triggered by external resources associated to the process, not including human resources). The process participants, whenever they realize the task goals have diverged from the operating conditions, directly trigger the manual detection. The Detection Service is responsible for documenting the triggering events in a database and invoking the Diagnosis Service.

Figure 4 – Detection Service.

The Diagnosis Service (Figure 5) collects assessment data related with a hazard. The data is obtained from the operators and whenever feasible from the BPM system. The following assessment data is solicited by this component (Mourão & Antunes, 2005):

- Affected processes and instances – list of processes and instances running in the BPM system that may have been affected;
- Affected persons – the persons that may have been affected by the hazard;
- Type of hazard – initially specified by the Detection Service, it may later on be redefined by the operators;
- Type of detection – automatic or manual, as specified by the Detection Service;
- Reaction time – if there is time to plan the recovery or not.
After diagnosis, the service will determine the type of human involvement, using a decision tree to consider: that none is necessary, in the case of prescriptive/planned situations; model adaptations should be accomplished by system developers, in the case of mixed/planned situations; and restricted ad-hoc/unstructured interventions will be necessary, in the case of unplanned situations. The distinction between planned and unplanned situations is determined by the specified reaction time. The distinction between prescriptive and mixed control is determined by the successive failure to apply failure handling and exception handling techniques. We note that diagnosis is an ongoing process, which means it may be repeatedly invoked to update the assessment data and re-evaluate human involvement. We also note the Diagnosis Service does not make the distinction between restricted ad-hoc changes and unstructured interventions, as the differences are related with model consistency managed by the Recovery Service.

After human involvement is considered necessary, the Diagnosis Service has to characterize it. This includes identifying the liable person, who is primarily responsible for the Diagnosis and Recovery Services. If the hazard was manually detected, then the person that triggered the event is the liable person. In the case of an automatic detection, the liable person is the person most directly involved in the affected process (for instance, the person who was responsible for a failed activity). The Diagnosis Service will also automatically determine the affected process participants, identifying all the persons responsible for the affected activities, information that is usually available in the BPM system. The problem and solution descriptions are short textual descriptions about what occurred and what should be done to resolve the situation, as perceived by the liable person. The Diagnosis Service offers the liable person the capacity to designate another liable person. It also allows the liable person to analyze the hazard trajectory, i.e. the timeline of diagnosis information and recovery actions.

After determining and characterizing human involvement, the Diagnosis Service invokes the Recovery Service (Figure 6). The Recovery Service is basically an interface to the BPM system capable to apply a set of quasi-atomic actions in process instances, such as cancel, jump forward and backward, repeat and suspend (Reichert, et al., 2003). The Recovery Service also implements/interfaces with the failure and exception handling techniques that may be implemented. And finally, the Recovery Service is also responsible for enforcing model consistency checks whenever the types of interventions in the BPM system correspond to restricted ad-hoc changes.
Failure and exception handling are handled by the BPM system and registered in the repository by the Recovery Service with the purpose to track events.

The final service we consider in resilient BPM is the Escalation Service. The main purpose of the Escalation Service is to support unstructured interventions in the BPM system, offering in particular map guidance (Suchman, 1987). As we have seen previously, this also involves collaboration support.

The Escalation Service is fundamental to bring more people to the recovery process. We may consider four levels of escalating human involvement:

- Horizontally moving the liability to another operator with no further contributions from the originator;
- Involving peers, when multiple co-workers may communicate with the liable person to help analyzing and discussing the problem;
- Vertically moving the liability to a supervisor, although allowing the worker to contribute to analyze and discuss the problem; and
- The group, when the liable person designates a group to get concurrently involved in the recovery actions.

The liable person is the only one able to escalate the diagnosis/recovery by involving another operator, peer, supervisor or group. We note that escalation is a dynamic process. The responsibility may reside in one person and in a while escalates to another person, peers, supervisor or group (Mourão, 2008).

The Escalation Service must be complemented with collaboration support, necessary to establish communication channels between the various persons participating in the process. External tools may implement the collaboration support, being the Escalation Service responsible for interfacing with the external tools whenever necessary and maintaining the exchanged information in the repository for later examination. In our implementations we have been using synchronous and asynchronous collaboration tools, including e-mail, text messaging and chatting. This approach also supports linking the Escalation Service with more complex group support and decision support systems, or even specialized emergency management tools (Sapateiro & Antunes 2009). The Escalation Service invokes collaboration for the persons selected by the liable person and may export the information available in the repository.

And finally the Escalation Service is also capable to instantiate ad-hoc tasks dedicated to monitor the BPM system evolution or other environmental conditions. The tasks themselves are managed by the BPM system. This functionality might be viewed as a tiny process model specifically dedicated to collect data about the BPM system evolution towards the resolution of the ongoing situation.
5.1. Information Model

Having described the fundamental services associated with resilient BPM, we complement the discussion showing the corresponding data models. Every BPM system is by definition an information system, developed around a specific data model. In spite of the differences in BPM data models, we will attempt to describe the generic data model necessary to implement resilient BPM. We nevertheless refer that our data model may be biased by having implemented the described services in the OpenSymphony platform (OpenSymphony).
Figure 9 – Detection model, organized around hazard trajectories. A trajectory serves to organize and maintain the evolution of events according to chronological time. Note that hazards may be associated with specific process instances and activities.

Figure 10 – Diagnosis model. These data elements support the decision tree that leads to determining human involvement in the recovery process. The profile history serves to maintain a chronological record of diagnoses.
6. Discussion

The major research topic addressed by this paper is extending BPM with the capacity to overcome hazards in business operations. Our approach to resilient BPM was inspired by HRO and especially their capacity to deal with various levels of severity, from simple failures to catastrophic accidents, based on situation awareness, knowledge representation, and flexible operations.

We propose a resilience framework based on two major criteria: control and response. We see control as an essential driver of flexibility, emerging from the Suchman’s observation that even in office work some latitude of decision should be put in the hands of the operators (Suchman, 1987). Many other researchers concur that human intervention is necessary when the system is not capable to coordinate the organizational activities and the reality diverges from modeled processes (Abbott & Sarin, 1994; Sheth, et al., 1996). We considered three levels of control: prescriptive, mixed and discretionary, which cover a large spectrum of possibilities ranging from strict technology-driven control to completely human discretionary control, passing by mixed situations where control may swing between humans and technology. It is the capacity to support discretionary control that affords sensitivity to operations and decision-making.

But we emphasize the support to discretionary control is somewhat in conflict with the traditional main goal of BPM systems: relinquishing control from the operators with the intentions to make the organizational behavior more predictable and optimize the operations. Therefore one important challenge the BPM developers must consider is how to reconcile prescriptive and discretionary controls, considering they imply very different behaviors and are based on quite different assumptions.

The second criteria considered in our resilience framework concerns response. Departing from the Perrow’s (1999) view that operations are unpredictable and inevitably lead to hazards, we make the distinction between two major response strategies, one considering there is a plan defining the model changes necessary to adapt the BPM system to a new organizational behavior, and another considering there is no such plan.

Developers face again an interesting challenge, since BPM naturally favors advance planning but organizations often face many situations where plans are not available (Saastamoinen, 1995) or planning is not even possible for lack of time. The challenge is effectively supporting dynamic and often unstable changes in
the BPM system while at the same time preserving the capacity to make sense of the events and lead the system towards normalcy after abnormal conditions.

Based on this framework, we reviewed the different techniques developed in the BPM field to cope with hazards. We identified five major categories: failure handling, exception handling, model adaptations, restricted ad-hoc changes and unstructured interventions. These categories offer incremental resilience. It is interesting to note that each one of these categories has intrinsic limitations. For instance, failure handling is capable to resolve many technological glitches, such as network and database failures, but does not support more complex solutions requiring human assessment. Quite the contrary, the unstructured interventions support human decision-making and action, but do not seem adequate to resolve minor technological glitches. Thus system developers should definitely consider the need to implement the whole collection of techniques and articulate their functionality in a coherent way.

Having deducted the need to integrate the various resilience modes, we developed and implemented an integration strategy. The adopted strategy is based on the following major considerations:

- At a given moment, we may regard the resilience framework to find out what are the most adequate techniques to handle a hazard. However, hazards often have a dynamic trajectory. They may start as a minor glitch, such as an activity failure, to later on enfold into a major organization-wide accident. Consequently, the integration strategy should be based on the dynamic diagnosis of hazards and associated contexts.
- The unstructured interventions are characterized by the capacity to intervene in the system independently of any constraints necessary to preserve model consistency. This means the users are entitled to lead the system to an unstable state. As previously discussed, under these conditions the system should offer “map” guidance, i.e. provide situation awareness about the hazard trajectory, the operators’ assessments along the hazard trajectory, and the actions taken to resolve, mitigate or contain its consequences. Thus the integration strategy also requires a coherent management of context information, effectively substituting control with awareness.
- Most of the current techniques supporting unstructured interventions do so in conjunction with collaboration support. Collaboration support is perceived as fundamental to diagnose complex situations and make decisions under incomplete information. Therefore, the integration strategy should also address collaboration support.
- BPM fundamentally deals with information. Therefore we may conceive extending the BPM data elements with additional elements necessary to manage the integration of the various recovery techniques, considering in particular the hazard trajectory, situation awareness and collaboration support.

Our implementation is based on four resilience services: detection, diagnosis, recovery and escalation. The Detection service is responsible for interfacing with the system components and human operators with the purpose to detect hazards. Manual and automatic detection have been considered. The diagnosis service is responsible for determining the best approach to handle a hazard. Most of the functionality of this service is associated with the situation where human intervention is necessary. In these cases, the diagnosis service allows designated operators characterizing who will be involved or may be affected in the handling process, and what is the problem and possible solution. Since hazards have a trajectory, diagnosis is a dynamic process running in parallel to that trajectory. One important feature of this service is maintaining an historical record of the hazard trajectory, thus facilitating situation awareness and retrospective analysis.

The recovery service manages the system-level interventions necessary to recover from hazards. Basically, this service interfaces with the BPM system components responsible for managing processes and activities. The escalation service manages the operators involved in the recovery process and the necessary collaboration. Regarding the type of engagement, we considered four alternatives, including co-workers, peers, supervisors and groups. This allows aligning human intervention with the most typical structures found in organizational behavior. Furthermore, this service also considers the interface with external collaboration tools according with two collaboration modes, synchronous and asynchronous, which again cover the most common collaborations tools currently adopted by organizations, including e-mail and instant messaging. The combination of various collaboration channels and communication modes affords tailoring the human intervention to the particular demands of the hazard and organization structure.
We also presented several data models necessary to implement the described resilience services. These data models are based on our BPM implementation using OpenSymphony, but should be easily reused in other BPM systems developed around relational databases (van der Aals & van Hee, 2002). As we illustrate in Figure 8, a small number of process management data elements is necessary to interface with the described services. More implementation details, as well as information regarding the evaluation of the proposed approach are published elsewhere (Mourão, 2008; Mourão & Antunes, 2004, 2005, 2007). What we should emphasize is that developing resilience services requires accessing the BPM static (data) and dynamic (functions) elements as an open box, for instance to implement event triggers and quasi-atomic actions. The user-interfaces necessary to interact with the resilience services, for instance instantiating monitoring tasks and manually detecting hazards, should also be integrated with the worklist handlers supplied by BPM technology. This way utilizing resilience services becomes a “normal” operation.

We would like to point out some issues we have not yet researched, which may challenge future implementations of the proposed approach. The dynamic interactions between hazard trajectory and recovery actions may become so complex that the operators may find themselves unable to recover the system. This may lead the system to instability and ultimately to a crash. This is naturally a consequence of widening too much the control spectrum to discretionary actions. Additional functionality could mitigate this problem, for instance supporting undo/redo of recovery actions and offering visualization tools to better understand the hazard trajectory and overall system status.

Another issue yet to be disentangled concerns the duration of recovery processes and their interaction with new hazards. The developed data models consider some dynamic aspects related with hazard trajectory, e.g. the severity may evolve, the diagnosis may also evolve and therefore recovery actions and escalation change over time. However, considering particular contexts where the recovery actions may take a long time to bring back the system to normal operations, maybe in the order of weeks or months, we should also consider the possible occurrence of new hazards interfering with the previously unresolved ones. This type of interference has not been modeled and tested in our implementation.

Also, the dynamic interaction between recovery and collaboration may lead to concurrent recovery actions commanded by several operators. Currently, we assume that under these circumstances the operators will use collaboration tools to discuss a tactic and coordinate themselves. However, as the number of participants increase, such an approach becomes less feasible. A more sophisticated approach would require more integration between collaboration tools and hazard recovery.

We should also consider that the cross-organizational context might bring new challenges to resilience support. One particular problem is that responsibilities and cause-effect relationships are spread throughout different systems, which may not be completely open to each other (Luo, et al., 2003). The supported collaboration modes may have also to be extended to cover cross-organizational collaboration.

And finally, it is important to recognize the proposed framework requires further evaluation, extending beyond technical feasibility. Such evaluation actions must be carried on different organizations and through long time periods, where data regarding reactions to hazards should be collected and evaluated to understand if the proposed solution addresses the increasing resilience capacity. This is necessarily a long-term project that should be carried out by our research group. Nevertheless, we are already preparing an implementation on a Portuguese organization capable to support these ambitious goals.

7. Conclusions

In this paper we construct a view over resilient BPM integrating NAT and HRO perspectives, the former considering that accident management has become normal organizational behavior and the later focusing on the need to develop safety, sensitivity to operations and decision-making functionality within socio-technical systems.

The literature review shows that resilience support is implemented with incremental measures, aiming to recover from hazards at different resilience levels, ranging from failure handling, exception handling, model adaptation, restricted ad-hoc changes, and unstructured interventions. One resource becoming particularly critical along this incremental path is the human. Of the five strategies, three of them involve humans: model adaptations require system developers to analyze and deploy new work processes at the blunt end (Woods & Hollnagel, 2006); restricted ad-hoc changes are accomplished by process participants under
technology control; and unstructured interventions are done by system operators at the sharp end using any available technology guidance.

We then characterized four core mechanisms considered necessary to implement resilient BPM: hazard detection, diagnosis, recovery and escalation. The escalation service is certainly the most distinctive service, being responsible for orchestrating the several operators necessary to overcome hazards using unstructured interventions. The escalation service supports the participation and collaboration necessary to build situation awareness and make decisions over what actions are necessary to overcome large-scale hazards.

The major outcomes of this research are:

- A framework characterizing resilient BPM in two dimensions: control and response;
- A review of the major techniques developed in the BPM field to overcome hazards, organized according the framework;
- Characterization of the fundamental advantages and drawbacks of unstructured interventions;
- A collection of services and associated functionality necessary to integrate all resilience levels in BPM systems;
- Characterization of the data models necessary to implement resilience in BPM systems developed around relational databases;
- In a more theoretical perspective, we also highlight the fundamental challenges and tradeoffs brought by resilient BPM: articulating prescriptive, mixed and discretionary control, reconciling dynamic changes with normal operations, and integrating discretionary control with situation awareness.

**Acknowledgements**

This research was partially funded by the Portuguese Foundation for Science and Technology, Projects PTDC/EIA 102875/2008 and 67589/2006.

**References**


