Business Process Risk Management and Simulation Modelling for Digital Audio-Visual Media Preservation

Vegard Engen, Galina Veres, Simon Crowle, Paul Walland
IT Innovation Centre, University of Southampton
Southampton, United Kingdom
Email: {ve, gvv, sgc, pww}@it-innovation.soton.ac.uk

Christoph Bauer
Multimedia Archives, Austrian Broadcasting Corporation
Vienna, Austria
Email: christoph.bauer@orf.at

***Abstract*—Digitised and born-digital Audio-Visual (AV) content presents new challenges for preservation and Quality Assurance (QA) to ensure that cultural heritage is accessible for the long term. Digital archives have developed strategies for avoiding, mitigating and recovering from digital AV loss using IT-based systems, involving QA tools before ingesting files into the archive and utilising file-based replication to repair files that may be damaged while in the archive. However, while existing strategies are effective for addressing issues related to media degradation, issues such as format obsolescence and failures in processes and people pose significant risk to the long-term value of digital AV content. We present a Business Process Risk management framework (BPRisk) designed to support preservation experts in managing risks to long-term digital media preservation. This framework combines workflow and risk specification within a single risk management process designed to support continual improvement of workflows. A semantic model has been developed that allows the framework to incorporate expert knowledge from both preservation and security experts in order to intelligently aid workflow designers in creating and optimising workflows. The framework also provides workflow simulation functionality, allowing users to a) understand the key vulnerabilities in the workflows, b) target investments to address those vulnerabilities, and c) minimise the economic consequences of risks. The application of the BPRisk framework is demonstrated on a use case with the Austrian Broadcasting Corporation (ORF), discussing simulation results and an evaluation against the outcomes of executing the planned workflow.**

Keywords–Risk management; business processes; workflows; semantic modelling; simulation modelling.

I. INTRODUCTION

Digital preservation aims to ensure that cultural heritage is accessible for the long term. From the 20th century onwards, AV content has provided a significant record of cultural heritage, and increasing volumes of AV content that have been digitised from analogue sources or produced digitally present new preservation challenges. The focus is no longer on reducing damage to the physical carrier by maintaining a suitable environment; rather, archives must ensure that the significant characteristics of the content, represented digitally, are not lost over time. Digital data enables easier transfer, copying, processing and manipulation of AV content, which is at once a boon but also a problem that requires continuous and active management of the data.

Digital damage is defined here as any degradation of the value of the AV content with respect to its intended use by a designated community that arises from the process of ingesting, storing, migrating, transferring or accessing the content. The focus here is on strategies that can be used to minimise the risk of loss. In particular, we focus on dealing with issues resulting from system errors, rather than random failure or corruption, considering the risks to the AV content as it is being manipulated by various activities in a workflow process. This includes risks introduced by the people, systems and processes put in place to keep the content safe in the first place.

Archival processes dealing with digital AV content are underpinned by IT systems. In the few years that archives have been working with digitised and born-digital content, best practice in terms of digital content management has rapidly evolved. Strategies for avoiding, reducing and recovering from digital damage have been developed and focus on improving the robustness of technology, people and processes. These include strategies to maintain integrity, improve format resilience and interoperability, and to combat format obsolescence.

This paper builds on [1], presenting the research and development work of a Business Process Risk management framework (BPRisk) developed in the EC FP7 DAVID project [2], which combines risk management with workflow specification. BPRisk has been designed to support a best practice approach to risk management of digital AV processes (and thus the content itself). In this paper, we will give an overview of this framework, focusing on semantic modelling, risk specification and simulation modelling. Within the DAVID project, this research and development has been conducted to provide a tool to help prevent damage to digital AV content in broadcasting archives, although the approach is clearly applicable to any digital archive management process where the same challenges of workflow and migration risk are present.

The BPRisk framework is generic in nature, supporting risk specification for Business Process Modelling Notation (BPMN) 2.0 [3] workflows in any domain. The framework utilises a novel semantic risk model developed in the project that encapsulates domain knowledge generated in the DAVID project on known risks (and controls) associated with activities in a controlled vocabulary for the domain of digital preservation (also developed in the project). This enables the framework to be an effective support tool to users who are typically not familiar with formal risk management. The semantic risk modelling provides the domain experts with

a starting point for conducting risk analysis, and semantic reasoning is utilised to provide suggestions of relevant risks and controls for the activities in the respective workflows at design time.

Another focus of this paper is the simulation modelling adopted in the BPRisk framework. The purpose of the simulation modelling is to help an organisation reduce costs by designing or optimising workflows in order to reduce the likelihood or impact of risks occurring. For example, it could be used to help justify expenses on technology and control tools, showing the anticipated cost of dealing with issues (risks) when they are not addressed (controlled) versus the cost of preventing them. That is, if the cost of prevention is less, one could argue an anticipated Return On Investment (ROI). Moreover, the simulations can help identify the key vulnerabilities in a workflow in order to help target investments. The aim is to expose issues at design-time before a workflow is actually executed.

In the DAVID project, the risk management work presented in this paper is one of the four cornerstones of interlinked work on i) understanding damage (how it occurs and its impact), ii) detecting and repairing damage, iii) improving the quality of digital AV content, and iv) preventing damage to digital AV content and ensuring its long-term preservation. The latter is a significant challenge, despite the advances in i) to iii), especially with respect to format obsolescence and failure in processes and people who handle the digital content, which is discussed further below.

The challenges and related work on digital preservation are discussed in Section II. Risk management in this domain is discussed in Section III. Thereafter, in Section IV, we present the BPRisk framework that has been developed in the DAVID project. Following this, we present and discuss further details of the semantic risk modelling and simulation modelling adopted in the framework in Section V and VI, respectively. Section VII discusses the application of BPRisk on a real use case with the Austrian Broadcasting Corporation. This includes simulation results from the planning stage of the workflow development, and a comparison with the outcomes from executing the workflow. Section VIII concludes this paper and discusses future work.

II. Digital Preservation

AV content is generated in vast quantities from different sources such as film, television and online media, environmental monitoring, corporate training, surveillance and call recording. There are many reasons why content needs to be retained and archived, which might be to enable content re-use for commercial, educational or historical purposes, but equally it might need to be retained and accessible due to regulatory compliance, for security or recording health and safety issues. Historically, the preservation of analogue content has been intrinsically linked to its method of production; specifically, the media that is used to carry the signal (the carrier). This means that archives preserved 'masters' on magnetic tape, film and even phonograph cylinders [4]. Where masters no longer exist or content was not professionally produced, archives needed to preserve 'access' copies on media such as vinyl records, VHS/Betamax tapes, and audio cassettes. To reduce the risk of damage, archives had to consider the physical characteristics of the media and care for the physical environment to which the media was sensitive (e.g., light, heat, humidity and dust) and to look after the machines that read the media. To increase the chances of being able to read the content again, archives often created copies of the artefact, in case one copy was damaged.

Nowadays, AV content is commonly born-digital and archives such as INA (the French national archive) and ORF (the Austrian broadcaster), who were partners in the DAVID project, have initiated digital migration projects to digitise the older, analogue, content [5]. Digital content (digitised or born digital) can be copied, transferred, shared and manipulated far more readily than its analogue equivalent. In a world of digital AV content, preservation is largely agnostic to the carrier that is used to store and deliver the content. Therefore, preservation and archiving is about making sure that the digital data is safe and that processes that manipulate the data do not cause damage. When referring to 'digital damage' in this paper, it is worth noting the following definition:

> *"Digital damage is any degradation of the value of the AV content with respect to its intended use by a designated community that arises from the process of ingesting, storing, migrating, transferring or accessing the content."* [5]

The above definition may seem broad. Indeed, it covers damage arising from failure of the equipment used to store and process digital content, as well as that arising from human error or from 'failure' of the process. The challenge for digital preservation is to keep the AV content usable for the long-term, which is threatened by format obsolescence, media degradation, and failures in the very people, processes and systems designed to keep this content safe and accessible [6], [7], [8].

Therefore, the core problem is greater than the potential for a digital file already in the archive to become damaged over time due to, e.g., bit rot [6], which can effectively be addressed by keeping multiple copies of each file [5], [7]. We also need to consider the future challenges for digital preservation as some analyses [9] predict that as ever more 8K AV content is ingested into archives, the growth in data volumes may outstrip the predicted growth in data capacity, but more importantly still, the data write speed necessary to store these high data volumes at real time will not be achievable, meaning that it will become impossible to cost-effectively store and replicate such content as it is produced. Therefore, strategies such as file-level replication may not be feasible in the future, and managing risk to the entire workflow process, and determining the most cost-effective archive management approach becomes essential.

III. Risk Management for Digital Preservation

Risk management, in a broad sense, can be understood as *"the coordinated activities to direct and control an organisation with respect to risk"* [10]. Risk, as defined by ISO 31000 [10], is the *"effect of uncertainty on objectives"*. In this context, *uncertainty* arises from random or systematic failure of preservation systems and processes (that may involve manual human activities). The *effect* of which is to cause damage to AV content. In general terms, we can say that the key *objective* is to ensure long-term preservation of digital AV content, i.e., avoid damage and ensure that it can be accessed in the future.

Current archives such as the French national archive (INA) and the Austrian Broadcasting Corporation (ORF) typically deploy a number of IT based strategies for avoiding, preventing or recovering from loss [5]. These archives are engaged in a process of long-term Digital Asset Management (DAM) [11], specifically Media Asset Management (MAM), which focuses on storing, cataloguing and retrieving digital AV content. Several commercial tools exist to support the MAM process, some of which support risk treatment strategies such as keeping multiple copies of each file (redundancy). However, these tools do not include a model of risk. The archive must decide on risk indicators and define the way in which these can be measured in order to monitor them, often using separate tools to do so.

Based on the analysis of threats to digital preservation in the DAVID project [7], [5], [12], it is clear that it is necessary to manage the threats to the workflow processes themselves. In this domain, we can note the following main sources of risk: equipment and tools (hardware, services/systems, software/algorithms), file formats (including implementations of standards and variations between versions), processes and human errors.

Workflows are often used to describe business processes and, increasingly often, are used to automate some or all of the process. Automated workflow execution is possible if the process is specified in a machine-interpretable fashion, such as using BPMN. In Hazard and Operability Studies (HAZOP), risks are seen as inherent in processes, as individual steps may fail, causing consequences for later parts of the process, or if the process is not executed correctly. Risk-aware business process management is critical for systems requiring high integrity, such as archives.

A recent review of business process modelling and risk management research has been conducted by Suriadi et al. [13], identifying three parts to risk-aware business process management:

- Static / design-time risk management: analyse risks and incorporate risk mitigation strategies into a business process model during design time (prior to execution).
- Run-time risk management: monitor the emergence of risks and apply risk mitigation actions during execution of the business process.
- Off-line risk management: identify risks from logs and other post-execution artefacts, such that the business process design can be improved.

Several approaches have been proposed to model business processes and risk information such that it enables risk analysis. Rosemann and zur Muehlen propose integrating process-related risks into business process management by extending Event-driven Process Chains (EPC) [14]. Risks are classified according to a taxonomy including structural, technological and organisational risks.

Analysis of process risks is difficult given that operational risks are highly dependent on the specific (and changing) business context. Many risks are caused by business decisions (e.g., preservation selection strategy or migration path), so large volumes of data required for statistical methods are often not available for analysis. Those who subscribe to this thesis use structural approaches, such as Bayesian networks, HAZOP and influence diagrams. For example, Sienou et al. [15] present a conceptual model of risk in an attempt to unify risk management and business process management using a visual modelling language.

In contrast to the above thesis, some believe that run-time analysis of risks is possible with a suitably instrumented execution process. Conforti et al. [16] propose a distributed sensor-based approach to monitor risk indicators at run time. Sensors are introduced into the business process at design time; historical as well as current process execution data is taken into account when defining the conditions that indicate that a risk is likely to occur. These data can be used for run-time risk management or off-line analysis.

Given that analysis of business processes using structured and/or statistical approaches can reveal vulnerabilities, it is important to control the risk that these vulnerabilities lead to loss. Bai et al. [17] use Petri nets (a transition graph used to represent distributed systems) and BPMN to model business processes and to optimise the deployment of controls, such that the economic consequences of errors (measured as Conditional Value at Risk - CVaR) are minimised.

Using BPMN, the PrestoPRIME project described the preservation workflows that were implemented in the preservation planning tool iModel [18]. It has shown that tools are required to model such generic preservation workflows in such a way that they can be related to specific preservation processes and augmented with information concerning risks.

IV. Business Process Risk Management Framework

Here, we present a Business Process Risk management framework (BPRisk) developed in the DAVID project (Section IV-C), designed to support the aims and risk management process discussed below in Sections IV-A and IV-B.

A. Aims of Risk Framework for Digital Preservation

Above, we have discussed the motivations for a risk management of business processes, according to the wider challenges in the domain of digital preservation. For digital preservation / archive management, the key actor we are addressing with the proposed risk framework is the preservation expert / specialist, who is responsible for designing workflows for managing and processing digital AV content. We can summarise here some key value-added aims of a risk management framework in the context of digital preservation:

1) Helping preservation experts develop new workflows, especially the early stages of development. Note that the purpose of the framework is not to replace MAM tools (discussed in Section III, above), nor the preservation experts, but to be a value-added tool to assist them.
2) Helping preservation experts optimise workflows (in terms of cost effectiveness and security), considering also trade-offs where too many corners are cut (to reduce cost), which may lead to increased risk.
3) Helping preservation experts communicate and justify decisions about choices for elements in workflows. This may be related to arguing expected financial ROI of putting in place certain risk mitigations, for

example. By risk mitigation, we here refer to reducing the likelihood or impact of risk.

4) Helping organisations change their processes, as the risk arising from such changes is typically seen as very high, which inhibits change. However, change is necessary to address the issue of format obsolescence.

From an organisational point of view, some of the key reasons to perform risk management can be summarised as follows:

1) Workflows can be large and complex. Therefore, there can be too many variables and options for preservation experts to consider simultaneously to accurately estimate the potential impact of risk.
2) Risk information is typically in experts' heads, which is itself a risk from the organisation's point of view. The risk framework ensures that the knowledge is captured and retained, and is readily available should the organisation be subject to an audit or the expert is unavailable or leaves the organisation.
3) Improve cost-benefit by a) identifying and understanding key vulnerabilities and b) targeting investments to address those vulnerabilities.
4) Move away from "firefighting". That is, organisations may spend more time dealing with issues rather than preventing them in the first place. Risk management is key to prevention, i.e., spending more time in the planning stages to save time and cost on dealing with issues in the future that could have been avoided.

It is important to note that the end users of the risk management framework in this context are unlikely to be risk experts. They are domain (preservation) experts, and they will be acutely aware of a wide range of potential issues concerning the preservation workflows they manage. However, the term risk and explicitly managing risk may be entirely unfamiliar and it is important that the risk management framework is suitably designed to aid the domain experts (rather than simply being a risk registry).

B. Risk Management Process

The risk framework should support a process that promotes best practices to address the aims discussed above in order to reduce the risks to long-term preservation. There is a natural focus on the planning aspects regarding risk management, but we do need to consider the wider context as well.

Several risk standards and methodologies exist, but it is not within the scope here to discuss them in detail. However, we will make reference to one in particular here, ISO 31000 [10], to show how it relates to a risk management approach proposed here based on the Deming cycle. The Deming cycle is a four-step iterative method commonly used for control and continuous improvement of processes and products. The four steps are: Plan, Do, Check and Act. For this reason it is also commonly referred to as the PDCA cycle, and is key to, for example, ITIL Continual Service Improvement [19]. In general terms, risk management is a part of continual improvement of processes – preservation workflows in this context.

The ISO 31000 [10] risk management methodology is depicted in Figure 1, below, which depicts the various stages from 'establishing the context' to 'treatment' (of risk) that is also cyclic. Supporting continual improvement of workflow processes is imperative in digital preservation, as discussed in Section II, as one of the key challenges in this domain is obsolescence and one of the key current risk strategies involving file-replication may not be feasible in the future.

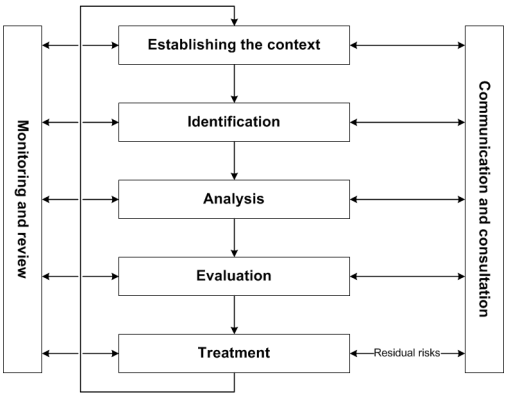


Figure 1. ISO 31000 risk management process.

Given the aims discussed above, each of the four stages of the Deming cycle is covered below from the perspective of what a user (preservation expert) would do, with reference to the related stages of the ISO 31000 methodology).

Plan ('establishing the context' and 'identification' stages of ISO 31000): build workflows, capture risk information, simulate workflow execution scenarios to identify key vulnerabilities and estimate impact of risk, and make decisions.

Do ('analysis' stage of ISO 31000): execute business process, orchestrate services, and record execution meta-data.

Check ('evaluation' stage of ISO 31000): analyse workflow execution meta-data and process analytics, calibrate simulations and trigger live alerts.

Act ('treatment' stage of ISO 31000 as well as feedback and loop-back to the previous stages): adapt workflows and manage risk. Re-run simulations (Plan), enacting the offline changes in the real business process and continues execution (Do) and monitoring (Check).

Note also how this relates to the three risk-aware business processes discussed above from Suriadi et al. [13]; static/design-time risk management (Plan), run-time risk management (Do) and off-line risk management (Check). The final step in the Deming cycle, Act, covers multiple processes.

C. Risk Components

Based on the above aims, a high level component view of the BPRisk framework developed in the DAVID project is depicted in Figure 2. This framework integrates both new components developed in the DAVID project as well as existing open source technologies, which is discussed below.

BPRisk Dashboard: The main entry point for the user from which the user can access the functionalities of the framework, e.g., to create workflows, specify risks, run and view risk simulation results, etc. Figure 2 also shows two vocabularies used, one for known domain-specific risk and one for domain specific activities. This is discussed further below.

Workflow Designer: There are several existing, mature, tools for this, supporting the well-known BPMN 2.0 standard, such as Signavio Decision Manager [20] and the jBPM

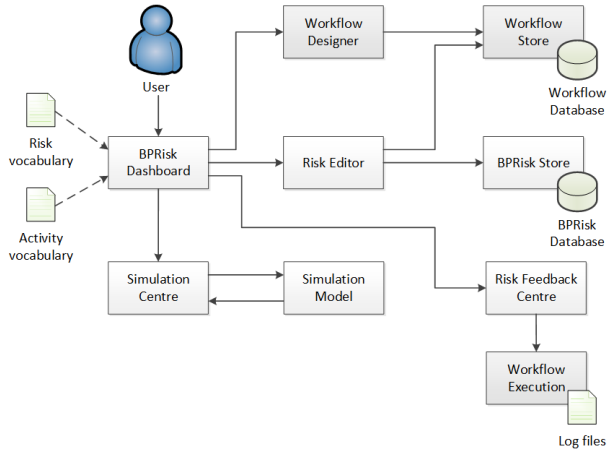


Figure 2. BPRisk framework high level component view.

Designer [21]. The latter has been adopted in the BPRisk framework as it is available as open source.

Workflow Store: This is a component to persist any workflows created, updated or imported. Existing tools, such as jBPM come with multiple persistence options and a RESTful API for accessing and managing the workflows.

Risk Editor: As described above, this component is responsible for allowing users to specify risks. As discussed earlier in this paper, the end-users of this system are not likely to be risk experts. Therefore, the Risk Editor utilises the two vocabularies mentioned above in a semantic risk model, which is used to aid users in specifying risks. See Section V for further discussion.

BPRisk Store: This is a component for persisting risk specifications and risk simulation results (a connection from the Simulation Centre has not been depicted in Figure 2 for the sake of simplifying the diagram).

Simulation Centre: This is a component for managing the running of simulation models for workflows annotated with risk information. This component deals with configuring different simulation scenarios and allows users to visualise and compare the results.

Simulation Model: A stochastic risk simulation model that the Simulation Centre can execute. This component simulates executions of the workflow process and the occurrences of risks defined for the workflow activities. As output, the simulation model gives information on, for example, risk occurrences, time and cost spent on risk, and impact of risk.

Risk Feedback Centre: A component for getting data from real workflow executions that can be used to a) analyse the workflow execution meta-data and b) to modify/adapt/calibrate the workflows (e.g., risk details) and simulation configurations to improve the accuracy for future simulation scenarios.

Workflow Execution: An external software component to the BPRisk framework, which would be invoked to execute a workflow process. This is a source of workflow execution data for the Risk Feedback Centre.

D. Implementation and Integration

A BPRisk framework prototype has been implemented as a RESTful [22] web application using Java Spring [23]. Figure 3 shows an architecture diagram of the key components that are in scope of this paper.

Web services are denoted with [WAR], comprising the BPRisk web application itself, a simulation service, the jBPM Designer (with Guvnor for workflow storage) and a Sesame service for the OWLim triple store used. The BPRisk web application follows a Model-View-Controller (MVC) [24] design pattern, comprising a shared data model (not discussed here), a control layer (blue) and view components for User Interface (UI) interactions with the users. Due to the MVC RESTful approach taken, it is possible for external client applications to access the data services, such as workflow data, in way that allows flexible compsition of relevant information for the end-user.

As noted above, the jBPM Designer has been integrated for workflow design, i.e., to graphically create new workflows or editing existing workflows. It supports the BPMN 2.0 standard, which allows users to specify workflows using, e.g., events (such as 'start' and 'end'), activities and connections between the activities, such as sequence flows or gateways to represent process logic. Figure 11 gives an example of a BPMN workflow using exclusive OR gateways.

The jBPM Designer uses the jBPM Guvnor as a workflow store by default, which has been adopted in BPRisk. However, to enable integration with other workflow management tools, the BPRisk framework has been designed such that workflow data is accessed via the Risk Data Service API, as depicted in Figure 3. More details are shown in Figure 4. A generic 'Workflow Accessor' component communicates with a specific workflow accessor module that functions as an adaptor. Here, a 'Guvnor Accessor' is shown, which will make a call to the Guvnor Service via its REST API in order to manage workflow information. Within the Risk Data Service API, the workflow data from Guvnor is processed via an 'Activiti BPMN parser' [25] before workflow information is returned in a shared BPRisk data model format (activities, gates and flows).

There are three different data storages depicted in the Risk Data Service API layer:

1) Workflow Store: for persisting and accessing the BPMN workflows, using the jBPM Guvnor as discussed above.
2) BPRisk Semantic Store: for storing semantic data pertaining to workflows, linking with controlled vocabularies for risks and domain specific activities, enabling semantic reasoning to aid users in creating or optimising their workflows.
3) BPRisk Project Store: for storing all other BPRisk data, including workflow projects, users, simulation configurations and simulation results.

A risk simulation model has been implemented in Matlab Simulink [26]. This has been integrated via a separate web service, to enable modularity and scalability as simulations can be computationally heavy. More details of the simulation modelling is provided in Section VI, followed by a discussion of simulation results in Section VII.

V. Semantic Risk Modelling

The BPRisk framework utilises a semantic risk model for specifying and reasoning about risks associated with workflow

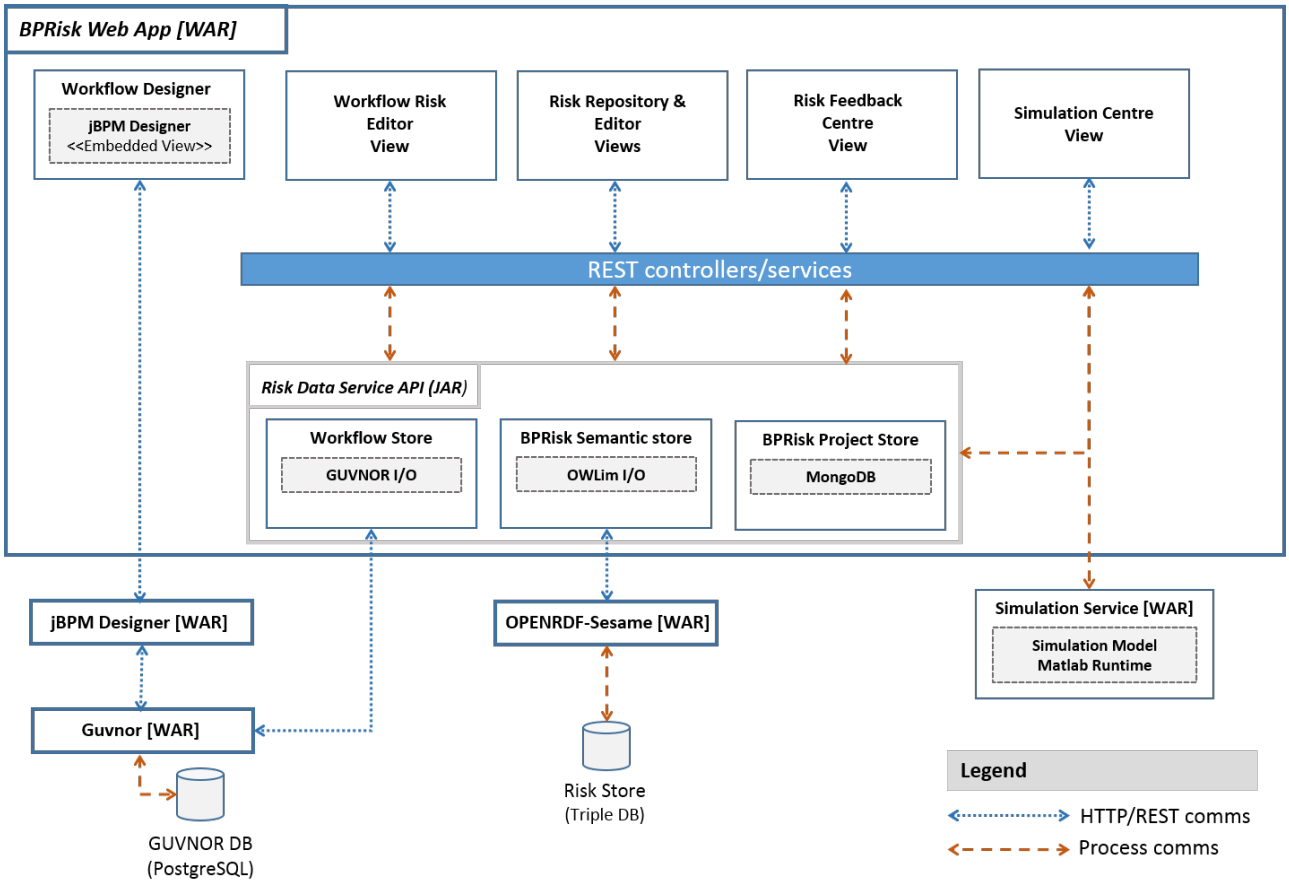


Figure 3. BPRisk architecture diagram.

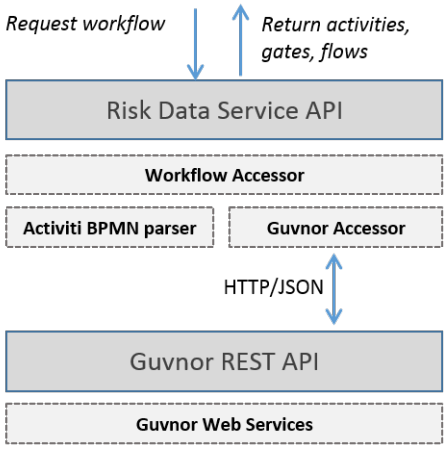


Figure 4. BPRisk workflow API.

activities. The modelling approach is generic in nature, utilising a multi-level ontology to include domain specific workflow activities and risks.

A. Modelling Approach

The BPRisk ontology represents information related to risks, controls and activities. This representation allows flexibility and extensibility of the risk model. It can be easily published (e.g., as a set of OWL files), and can be extended in unexpected ways. For example, the BPRisk ontology allows for the possibility of injecting provenance based information that can provide an auditable trail linking the identification of a risk factor (related to a workflow element) to its subsequent treatment using a provenance based ontology such as W3C PROV [27].

The approach to building the ontology is based on work done in the SERSCIS project [28]. The authors use a layered, class-based ontology model to represent knowledge about security threats, assets and controls. Each layer inherits from the layer above. The CORE layer describes the relationships between a central triad (threat, asset, control). A domain security expert creates sub-classes for each of these core concepts to create a GENERIC layer. A system expert further sub-classes the generic concepts to specialise them for the system of interest, creating the SYSTEM layer. Note that this ontology was used in the context of modelling systems and interactions between system components, where it is assumed that a system of a particular type is always subject to the threats identified by the security and system experts. This expert knowledge, therefore, helps the system designer create more secure systems as they may not have this expert knowledge themselves.

The same, layered, ontological approach has been taken

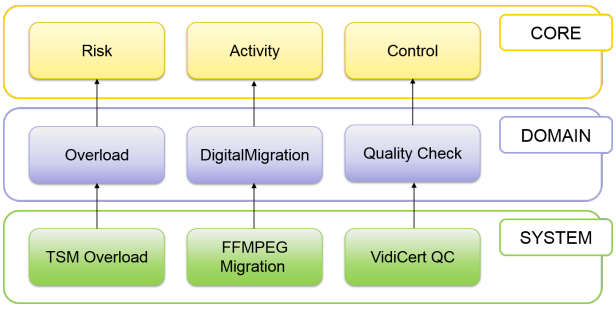


Figure 5. Workflow risk ontology layers.

here, as illustrated in Figure 5, though with a few modifications. While the triad in the CORE layer in SERSCIS includes *Asset*, there is only one asset of value in this context – the digital AV object, which can be affected by different *Activities* in a workflow process (e.g., ingest, storage and transcoding). The term *Threat* used in SERSCIS can be understood as *Risk* in this context. Therefore, the CORE layer in BPRisk comprises a triad of *Risk*, *Activity* and *Control*.

The GENERIC layer from SERSCIS has been renamed to the DOMAIN layer here, as it better reflects the level at which knowledge of domain specific (generic) activities and risks are represented. It is at this level, we incorporate controlled domain vocabularies, which are discussed further below in Section V-C. This layer can be further extended via the SYSTEM layer by users of the BPRisk application.

B. Model Definition

The model focuses on the *Activities* in the preservation life cycle and the *Risks* that are inherent in their execution. *Controls* can be put in place to block or mitigate these *Risks*. The CORE layer comprises *Risk*, *Activity* and *Control*, as well as basic relationships such as 'Risk *threatens* Activity' and 'Control *protects* Activity'. However, the relationship between *Control* and *Risk* is established via rules that abstractly encode how types or super-types of both controls and risks can be linked together (see the following section), to determine the appropriate relationship. That is, a *Risk* is only considered *Mitigated* if an appropriate *Control* is in place. This is illustrated below in Figure 6.

The DOMAIN layer has been developed in the DAVID project for digital preservation, which describes common preservation activities, risks and controls. These are modelled as sub-classes, which can be quite hierarchical. As an example, the DOMAIN level classes in Figure 6 include three sub-classed *Activities*, 'Migration', 'Digital Migration' and 'Transcoding', with an associated risk 'Migration Fails'. Migration in this context refers to converting content in one format into another format. Digital migration refers to converting older analogue content into digital form.

The SYSTEM layer is a further extensible part that would be populated by the users of the BPRisk framework when they build a workflow of specific *Activities* and associate *Risks* to them. For example, a migration workflow may use a specific *transcoding* tool such as FFmpeg [29], which may have specific technical risks not covered by the more generic 'Transcoding' activity. Thus, a 'FFmpeg Transcoding' activity

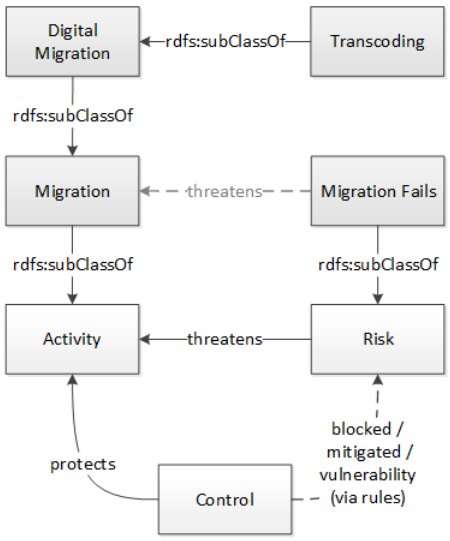


Figure 6. BPRisk ontology with sub-classing examples. CORE layer entities depicted in white and DOMAIN layer entities in grey.

may be added as a sub-class of 'Transcoding' (see Figure 6). This sub-classing is important, as we can reason about risks throughout the hierarchy, as discussed further in Section V-D.

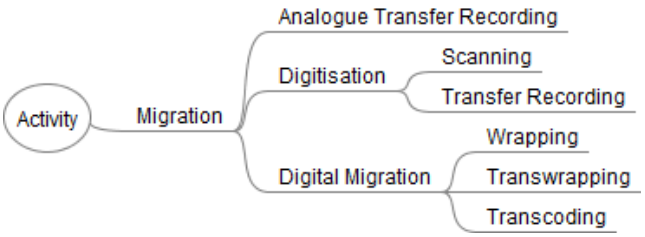
C. Controlled Vocabularies

In order to enhance the usability of the BPRisk framework, expert domain knowledge is included via the DOMAIN layer of the ontology presented above in Section V-B. The domain knowledge is incorporated via two controlled vocabularies: 1) known domain activities; 2) known risks and controls to the aforementioned activities.

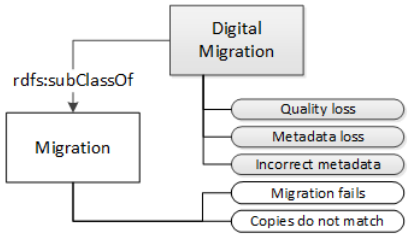
Although the controlled vocabularies reside within one of the three logical layers in the BPRisk ontology, there can be an extensive hierarchy of entries, which is depicted in Figure 7a. For example, we can see that 'Transcoding' is a type of 'Digital Migration' which is a type of 'Migration' activity. Further, for each activity, the controlled risk vocabulary defines common risks and controls for the known activities, such as for the 'Digital Migration' activity, as depicted in Figure 7b. For further details of the activities, risks and controls in the controlled vocabularies, interested readers are referred to [5].

When a user defines a workflow in the BPRisk framework, the domain knowledge embedded in the semantic model is used as follows. First, the activities from the BPMN workflow are extracted via the Risk Data Service API (see Figure 4). The user then maps each BPMN activity to activities in the BPRisk ontology (using the controlled activity vocabulary). Following this, the user will retrieve suggestions of potential risks, as per the controlled risk vocabulary. For example, for a 'Digital Migration' activity, the users will be presented with five possible risks, two of which are inherited from the parent activity 'Migration', as depicted in Figure 7b. The users can then chose which risks apply to the specific activity in the respective workflow.

As noted above, users are also able to add new risks, activities and controls not reflected in the DOMAIN layer. These user-specific additions form part of the SYSTEM layer,



(a) Hierarchy of Migration activities.



(b) Risks for the Digital Migration activity.

Figure 7. Examples from controlled vocabularies.

extending the knowledge repository. This knowledge is then available to users when working on other workflows. In the following section, we delve into further details on how this functionality is achieved via semantic reasoning.

D. Semantic Reasoning

Providing expert knowledge to users of the BPRisk framework is achieved via semantic reasoning, which is performed according to pre-defined rules.

Rule Composition: The layered abstractions CORE, DOMAIN and SYSTEM in the BPRisk model provide a useful framework within which to characterise generally increasing levels of specialism with respect to workflow activities and their associated risks and controls. Aligned with this arrangement, the rules that create relationships between activities, risks and controls are also expressible within and between these layers. For example, during a video migration activity a risk exists that the copies (that is, the migrated video) will not match the source material. This simple rule is expressed as using the Turtle RDF formalism [30]:

```
:CopiesDoNotMatch a owl:Class ;
  rdfs:subClassOf core:Risk ;
  rdfs:subClassOf [ a owl:Restriction ;
      owl:onProperty core:threatens ;
      owl:someValuesFrom act:Migration
  ] .
```

This risk is applied generally in the DOMAIN level and also to activities derived from the *migration* type - these include *analogue transfer recording*; *digital migration* and *digitisation* activities (as seen in Figure 7a, above). Workflows that include activities that belong to (or are themselves specialities of) this type will automatically generate an instance of this risk when processed by the BPRisk framework. Having identified a risk, one or more controls should be put in place to manage its potential outcomes. Here, the knowledge encapsulated in BPRisk rules can also be used. Activities are said to be 'protected' by controls that are available to manage risk; one very simple protection against a copy mismatch during migration would be to first check the migration

```
:CheckMigration a owl:Class ;
  rdfs:subClassOf core:Control ;
  rdfs:subClassOf [ a owl:Restriction ;
      owl:onProperty core:protects ;
      owl:someValuesFrom act:Migration
  ] .
```

and second, re-do the migration, if required:

```
:RedoMigration a owl:Class ;
  rdfs:subClassOf core:Control ;
  rdfs:subClassOf [ a owl:Restriction ;
      owl:onProperty core:protects ;
      owl:someValuesFrom act:DigitalMigration
  ] ;
  rdfs:subClassOf [ a owl:Restriction ;
      owl:onProperty core:protects ;
      owl:someValuesFrom  act:Migration
  ] .
```

Both of these control measures would be suggested by the BPRisk framework when the *'copies do not match'* risk is detected. Sometimes more specific knowledge is available that enhances the more general control procedures offered by the framework. In our example, those specific to protecting digital migration activities would be suggested. Below we see that re-introducing missing meta-data is a possible solution for risks threatening digital migration.

```
:ReintroduceMissingMetadata a owl:Class ;
  rdfs:subClassOf core:Control ;
  rdfs:subClassOf [ a owl:Restriction ;
      owl:onProperty core:protects ;
      owl:someValuesFrom act:DigitalMigration
  ] .
```

When a risk has been identified and controls put in place they can be marked up as either *blocked* or *mitigated*.

Rule Encapsulation: Encapsulating the relationships between risks, controls and activities are ultimately encoded as risk classification rules within the ontology knowledge based itself, using SPIN [31]. From a technical point of view, this provides a more flexible method of extending and executing rules incurring zero or only minimal changes to the compiled source used to operate on the results. Running inferencing over the model automatically applies the classification and can also determine the revised state of a workflow when control procedures have been put in place. In our earlier example, we considered the application of controls to act in the presence of risks threatening migration activity. In the SPIN formalism below, we express the fact that the control *check migration* blocks the *copies do not match risk* that is generated in the presence of a *migration* activity.

```
:CopiesDoNotMatch_BlockedBy_CheckMigration_For_Migration
  rdf:type spin:ConstructTemplate ;
  spin:body [
      rdf:type sp:Construct ;
      sp:templates (
          [
            sp:object risk:BlockedRisk ;
            sp:predicate rdf:type ;
            sp:subject [ sp:varName "r"^^xsd:string ; ] ;
          ] ) ;
      sp:where (
        [
          sp:object act:Migration ;
          sp:predicate rdf:type ;
          sp:subject [ sp:varName "a"^^xsd:string ; ] ;
```

```
        ]
        [
          sp:object risk:CopiesDoNotMatch ;
          sp:predicate rdf:type ;
          sp:subject [ sp:varName "r"^^xsd:string ; ] ;
        ]
        [
          sp:object ctrl:CheckMigration ;
          sp:predicate rdf:type ;
          sp:subject [ sp:varName "c"^^xsd:string ; ] ;
        ]
        [
          sp:object [ sp:varName "a"^^xsd:string ; ] ;
          sp:predicate core:threatens ;
          sp:subject [ sp:varName "r"^^xsd:string ; ] ;
        ]
        [
          sp:object [ sp:varName "a"^^xsd:string ; ] ;
          sp:predicate core:protects ;
          sp:subject [ sp:varName "c"^^xsd:string ; ] ;
        ] ) ;
  ] ;
rdfs:subClassOf :RiskClassificationRules ; .
```

In running SPIN rules every time the knowledge about specific workflow activities (contained in the SYSTEM layer) is added we are able to automatically recognise, control and manage risks in a pro-active manner.

As noted above, the SYSTEM layer is developed so that it sub-classes the DOMAIN layer for a specific organisation using the BPRisk framework, as seen above in Figure 6. This should specify the kind of activity in the preservation workflow of interest, e.g., sub-class *Migration* as *DigitalMigration* as seen above in the examples from the DOMAIN layer. Workflow-specific risks can then be automatically generated; for example, the following is a generic construction rule to generate all risks:

```
CONSTRUCT {
    ?uri a owl:Class .
    ?uri rdfs:subClassOf ?gr .
    ?uri rdfs:subClassOf _:b0 .
    _:b0 a owl:Restriction .
    _:b0 owl:onProperty core:threatens .
    _:b0 owl:someValuesFrom ?sa .
} WHERE {
    ?sa (rdfs:subClassOf)+ act:Activity .
    ?sa rdfs:subClassOf ?ga .
    ?gr rdfs:subClassOf core:Risk .
    ?gr rdfs:subClassOf ?restriction1 .
    ?restriction1 owl:onProperty core:threatens .
    ?restriction1 owl:someValuesFrom ?ga .
    FILTER NOT EXISTS {
        ?uri rdfs:subClassOf _:0 .
    } .
    FILTER STRSTARTS(str(?sa),
       "http://david-preservation.eu/bprisk#") .
    BIND (fn:concat(STRAFTER(str(?gr), "#"),
       "_", STRAFTER(str(?sa), "#")) AS ?newclass) .
    BIND (URI(fn:concat(fn:concat(STRBEFORE(str(?sa),
       "#"), "#"), ?newclass)) AS ?uri) .
}
```

This rule finds all activities in the SYSTEM layer and creates a workflow-specific risk for each of the DOMAIN layer risks that threaten the activities' parent class. The name of the workflow-specific risk in this example is generated by concatenation of the DOMAIN layer risk name and the workflow-specific activity name.

E. Discussion

The purpose of the semantic modelling in the BPRisk framework, as mentioned earlier in this paper, is to support the end-users who are typically not risk experts in optimising and building more robust workflows in order to ensure the long-term value of their digital content.

Formal representation of domain knowledge (linking activities, risks and controls) using the BPRisk semantic framework confers upon its users the ability to encapsulate and operationalise expertise in the preservation of media in the context of workflow based processes. Knowledge is structured in terms of i) hierarchies that are capable of expressing general and specialisations of cases (activities or risks) and ii) bespoke networks of connected activities, risks and controls that combine to form rules that flexibly express scenarios applicable to a wide range of workflows. In the example provided above, we explore this ability in the scenario where a risk is mapped to the type (and sub-types) of migration. Here this specific risk is defined as relevant in the context of migration type activities and is managed by a particular recommended control. However, note that the same risk type may also be applicable to other unrelated activities but may not call for the same controls. Managing risk using this formalism, thus, offers the user customisable responses to risk depending on the activity in hand. This has been made possible through the use of an ontological approach to knowledge engineering in which RDF and SPIN technologies have been used to build a knowledge base that is readily extensible by end-users (via the BPRisk Dashboard user interface). This approach, therefore, adds value to media workflow assets by augmenting them with expertise that can be queried and refined at design time, then tested and updated (through simulation and feedback from real-world exectuion, as described in the following section).

A large part of the knowledge base described here is particular to *media workflow* risk management. However, the application of the BPRisk framework is not limited to this enterprise and could be applied to other problem domains in which risk within workflows play a significant role. The underlying BPRisk architecture and services would remain the same, but it should be noted that a significant initial effort would be required to capture and transform knowledge gathered from experts in order to populate the domain layer of the ontology with common activities, risks and controls.

VI. Simulation Modelling

In this section, we present the work done on workflow risk simulation modelling, which is an integral part of the BPRisk framework. In respective sections below, we discuss the purpose of the simulation modelling, the risk impact model, the risk generation model and risk control procedures. Thereafter, in Section VII, we will present results from simulation modelling on a workflow at ORF, the Austrian Broadcasting Corporation.

A. Purpose of Simulation

The purpose of risk simulation modelling is to help an organisation to reduce cost by designing or optimising workflows in order to reduce the likelihood or impact of risks occurring. For example, it could be used to help justify expenses on technology and quality control tools, showing the anticipated cost of dealing with issues (risks) when they are not addressed (controlled) versus the cost of preventing them. The costs may be less, so we can say there is a ROI. The simulations can help identify the key vulnerabilities in a workflow and to help target investments.

The aim is to expose issues at design-time before a workflow is actually executed. In this paper, the risk simulation

addresses issues that could occur in activities conducted in preservation workflows. There could be technical risks, such as a system operation failing, or human errors such as mistakes being done because the person is overloaded by too much content to deal with.

To this end, a stochastic risk management model was developed. This model allows users to simulate different scenarios and to produce confidence intervals for different risk measures, if required, by means of Monte Carlo simulations. Moreover, the stochastic model allows end-users to explore 'what if' scenarios and can be used both during planning and operation stages. The proposed stochastic risk management model consists of three main parts:

- Risk Impact Model.
- Risk Generation Model.
- Risk Control Procedures.

Below we describe each of these parts in detail.

B. Risk Impact Model

To classify possible risks (threats) in digital preservation, we have adopted the Simple Property-Oriented Threat model (SPOT) for Risk Assessment. The SPOT model [32] defines six essential properties of successful digital preservation: Availability, Identity, Persistence, Renderability, Understandability, and Authenticity. Interested readers are referred to the original article for details. However we will give a short definitions of each property and list threats associated with this property.

Availability is the property that a digital object is available for long-term use. *Threats*:

- A digital object deteriorated beyond restoration power.
- Only part of the digital object is available for preservation.
- A digital objects is not available for preservation due to disappearing, cannot be located or withheld.

Identity is the property of being referenceable. A limited amount of metadata is required for this property. *Threats*:

- Sufficient metadata is not captured or maintained.
- Linkages between the object and its metadata are not captured or maintained.
- Metadata is not available to users.

Persistence is the property that the bit sequences continue to exist in usable/processable state and are retrievable/processable from the stored media. *Threats*:

- Improper/negligent handling or storage.
- Useful life of storage medium is exceeded.
- Equipment necessary to read medium is unavailable.
- Malicious or/and Inadvertent damage to medium and/or bit sequence.

Renderability is the property that a digital object is able to be used in a way that retains the object's significant characteristics (content, context, appearance, and behaviour). *Threats*:

- An appropriate combination of hardware and software is not available, cannot be operated or maintained.
- The appropriate rendering environment is unknown.
- Verification that a rendering of an object retains significant characteristics of the original cannot be done (e.g., a repository is unable to perform sufficient quality assurance on migration due to volume).
- Object characteristics important to stakeholders are incorrectly identified and therefore not preserved.

Understandability requires associating enough supplementary information with archived digital content such that the content can be appropriately interpreted and understood by its intended users. *Threats*:

- The interest of one or more groups of intended users are not considered.
- Sufficient supplementary information for all groups of intended users is not obtained or archived.
- The entire representation network is not obtained or archived.
- Representation network of supplementary information is damaged or otherwise not renderable in whole or in part.

Authenticity is the property that that a digital object, either as a bitstream or in its rendered form, is what it purports to be. *Threats*:

- Metadata and/or documentation are not captured.
- Metadata maliciously or erroneously describes the object as something it is not.
- A digital object is altered during the period of archival retention (legitimately, maliciously or erroneously), and this change goes unrecorded.

Since not all possible threats/risks in digital preservation workflows will fall in the six properties mentioned above, we introduce an extra possible state in the SPOT model for such cases: ***Other***.

C. Risk Generation Model

The stochastic risk generation model is based on simulating a workflow in which risks associated with workflow activities take place based on risk occurrence probabilities and dependencies between risks. Dependencies between risks can be within a single activity or between consecutive activities. Below, we given an overview of the data that is needed for workflow simulation, divided into the following categories for convenience: general, workflow-related, risk-related, control-related and other simulation parameters.

General data:

- The purpose of the workflow under consideration. That is, what the workflow does, inputs and outputs to/from the workflow.
- The objectives of risk analysis for this workflow.

Workflow-related data:

- List of activities in the workflow, a short description of each activity, and how the activities are connected between each other.
- Decision points in the workflow, and based on records or previous experience how often each decision are

usually made at each decision point (e.g., at decision point 1, D1 will be made approximately 90% of the time and D2 10% of the time).

Risk-related data (for each activity):

- List of risks (threats) which can take place and their descriptions.
- Any dependencies between the risks in the same activity and/or risks from different activities. E.g., can the risks in the same activity occur simultaneously?
- Frequency of each risk occurrence, either from records or estimated based on the previous experience; frequencies of more than one risk taken place in a activity if relevant; any changes in frequency of occurrence of some risk in the activity if other risk in the same activity took place.
- For each risk
  - Probability (frequency) of occurrence.
  - Detection level (if known).
  - Negative impact on workflow measured in monetary values, percentages or some impact scale.
  - Affected SPOT properties (explained further below).
- If combination of risks can occur:
  - Frequency of combined occurrence.
  - Multiplication factor, which is used to update the probability of a risk if combinations of risks occur either in the same or previous activity.

Control-related data (for each risk):

- Is anything done on the fly (Ad-Hoc Control)? If yes, is the Ad-Hoc Control procedure covered by budget overheads? How effective is the Ad-Hoc Control procedure (a value for 'Expected Success' would be provided)?
- Are Active Control procedures available for a given risk and activity? If so, is there a delay before the Active Control takes place?
- List all other control procedures dealing with this risk and their effectiveness.
- If more than one procedure dealing with risk is available, describe conditions when different procedures are activated.
- List costs associated dealing with risk and time spent on dealing with risk.
- Describe how negative impact is reduced when Ad-Hoc or/and other control procedures are applied.

Other simulation parameters:

- Number of items to be processed through a workflow.
- Annual throughput of items.
- Number of items to be processed during a day, week, month or year.

The risk occurrence probability is calculated based on the Estimated Frequencies (EFs) of risks provided by a user. We assume here, that EFs of risks are based on a pre-defined annual throughput of a given workflow, and, therefore, a risk occurrence is simulated on a per item basis. If a pre-defined annual throughput of the workflow changes, the new estimated frequencies can be updated using Estimated Frequency Factor (EFF) provided by the user. For example, if we are interested in processing N items per month, which is equivalent to $12 \times N$ items per year, then the EF for risk i in activity j can be calculated as:

$$EF_{new}(i,j) = 12 \times N \times EFF(i,j) \tag{1}$$

where $EFF(i,j)$ is the Estimated Frequency Factor for risk i in activity j.

$EF_{new}(i,j)$ represents a frequency per year in this form.

The probability of risk occurrence based on EF per item can be calculated as follows.

$$P(risk\,for\,1\,item) = \frac{EF}{N_{items}} \tag{2}$$

where EF is the estimated frequency based on a pre-defined throughput for a given workflow in pD (per day), pW (per week), pM (per month) and pY (per year).

N_{items} corresponds to a number of items that can be processed in a day/week/month/year, based on this pre-defined throughput.

If the throughput of items is changed, then instead of EF, EF_{new} *will be used in conjunction with new values* pY, pM, pW *and* pD.

By the nature of dependency between risk occurrences, all risks can be divided in the following groups:

1) Risks do not have any known dependency between each other and it is assumed that they cannot occur simultaneously.
2) The frequency of one risk in a given activity changes temporally if another risk in this activity took place. In this case these risks can occur simultaneously. Otherwise, both of the risks can occur only one by one (mutually exclusive).
3) The frequency of ***Risk A*** in the next activity changes temporally if ***Risk B*** in the previous activity occurs. This situation will be modelled as follows: if ***Risk B*** took place, then the frequency of ***Risk A*** is changed accordingly (for a single run of the workflow).
4) The risks can occur simultaneously. In this case the frequencies of each risk and the frequency of risks occurring simultaneously are used to simulate such a situation. In this case 'Estimated Frequency' means that only a given risk took place, 'Frequency of Combined' shows the joint frequency of risks.

The identification of risk generation groups will be done automatically by checking corresponding values in order of priority:

1) Multiple Risk-Entry per Activity.
2) Multiplication Factor.
3) Frequency of Combined.

A Detection Level parameter allows us to simulate whether a risk was detected or not. If risk is detected, then procedure dealing with risk will be put in action (see the next section). Otherwise we mark affected SPOT properties and record level of Negative Consequences (NC). An example of risk specification and related simulation configuration is given in Table I.

D. Risk Control Procedures

The stochastic risk management model implements a procedure of dealing with risk that comprises two types of controls: **Ad-Hoc Control** and **Active Control**. These controls only apply if a risk is actually detected. The schematic presentation of Risk Control procedures is shown in Figure 8. If a risk is detected, then the Ad-Hoc Control procedure is started. Active Control is applied only if a cost spent on an Ad-Hoc Control procedure is higher than a pre-defined value. This is generally only likely to be issued for large and significant risks and could be, for example, re-training staff or allocating more resources. This type of control would typically incur additional cost and time to be put into place. However, note that Active Control is not necessarily available for all activities/risks. In this case only the Ad-Hoc Control procedure is applied to dealing with those risks.

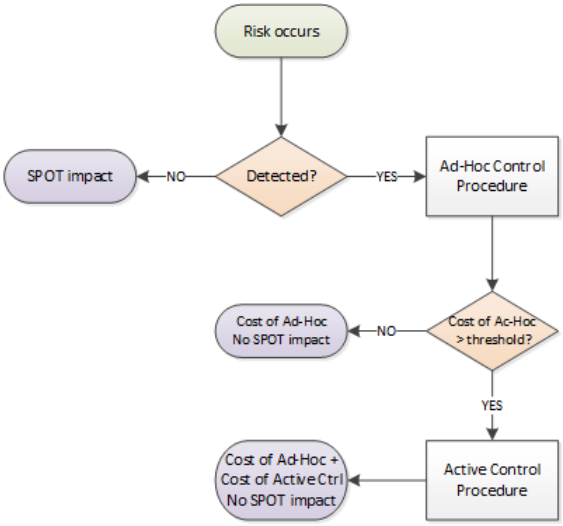


Figure 8. High level risk control flow chart. For simplicity, negative consequences are not shown, but do occur along with 'SPOT impact'. Similarly, time spent on dealing with risk is omitted; just referring to financial cost.

The details of the Ad-Hoc Control procedure are illustrated in Figure 9. It can be covered by overhead or not. Overhead is a term used here for either a budget or a percentage of resources set aside *a priori* to cover the cost of dealing with issues. If not covered by overheads, it will result in cost and time spent with dealing with the risk. However, if the Ad-Hoc Control procedure is covered by overhead and has 100% Expected Success of Ad-Hoc counter-measures, then a) the risk does not have any effects on the assets properties (SPOT model) or Negative Consequences and b) there are no (extra) costs associated with dealing with the risk.

If the Expected Success of the Ad-Hoc Control procedure is not 100%, then, based on the number of items to be processed

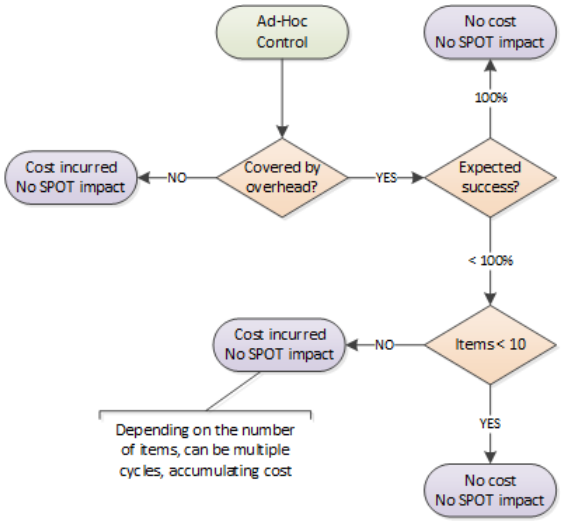


Figure 9. Ad-Hoc Control procedure flow chart. For simplicity, negative consequences are not shown, but do occur along with 'SPOT impact'. Similarly, time spent on dealing with risk is omitted; just referring to financial cost.

through the workflow and Excepted Success rate, additional Ad-Hoc Control procedures are performed as follows.

1) For up to 10 items: the 1st attempt of the Ad-Hoc Control procedure always successful independent of the Expected Success rate provided for the respective risk.
2) For 11 to 100 items: 2nd attempt will be always successful if the Expected Success rate is 50% or above. 3 attempts will have to be made otherwise to achieve success.
3) For 101 to 1,000 items: 3rd attempt will be needed to reduce the Negative Consequences to zero.

Note that, in this case, the Ad-Hoc Control procedure is performed until Negative Consequences is zero or near zero. In general, the number of attempts needed is equal to the order of the number of items to be processed via the workflow. Let us take an example:

A risk is detected for 500 items, and we have a 90% Expected Success rate for this risk.

This means the 1st attempt will resolve the issue for 450 items → 50 items remaining.

The remaining 50 items are subject to a 2nd attempt → 5 items remaining.

The remaining 5 items are then subject to a 3rd and final attempt, and for up to 10 items, we have modelled the attempt to have a 100% success rate, regardless of the value provided for the Expected Success rate.

For each attempt, time and cost is accumulated, and it is this sum that is subject to the Active Control check; i.e., if this exceeds some pre-defined threshold.

The general formula for calculating costs of the Ad-Hoc Control procedure is as follows:

TABLE I. Risk specification example.

	TSM Retrieve		**Mapping Control**	
	Wrong file selected	***Retrieve fails***	***Overload***	***Wrong assessment***
Estimated Frequency	5 per year	5 per year	2 per month	2 per month
Estimated Frequency Factor	0.0004166	0.0004166	0.002	0.001
Multiplication Factor	1	1	1	5
Multiple Risk Entry per Activity	No	No	Yes	Yes
Frequency of combined	None	None	None	None
Detection Level	90%	100%	100%	75%
Level of Severity	1	1	2	2
Expected Success of Ad-Hoc counter-measures	100%	90%	50%	90%
Cost associated with Risk (CAR per hour)	€50	€50	€70	€50
Time spent on dealing with risk (TAR per item)	0.1 hrs	0.2 hrs	0.5 hrs	0.2 hrs
Cost for Active Control strategy (CACS)	€800	None	€1,000	€800
Active Control Activation rule	$(CAR \times TAR) \times 1.2 > CACS$		$(CAR \times TAR) \times 1.0 > CACS$	$(CAR \times TAR) \times 1.2 > CACS$
Delay of Active Control (days)	5	0	5	22
SPOT Availability	1	1	1	1
SPOT Identity	0	0	0	0
SPOT Persistence	0	0	0	0
SPOT Renderability	0	0	0	0
SPOT Understandability	0	0	0	0
SPOT Authenticity	1	0	1	1

$$cost = TAR \times CAR \times \sum_{i=1}^{k} n_i \tag{3}$$

where TAR is the time needed for dealing with risk for one affected item,

CAR is a cost associated with dealing with risk for 1 hour TAR.

k is a number of attempts of the Ad-Hoc Control procedure calculated as described above, based on a number of items passing through the workflow.

n_i is a number of affected items after each Ad-Hoc effort calculated as a product of the number of affected items before the i^{th} attempt and (1 - success rate of Ad-Hoc) in decimal points.

Active Control is applied only if a cost spent on Ad-Hoc Control exceeds a pre-defined allocation of available resources for Ad-Hoc Control (CACS). Active Control does not need to be defined for all activities/risks. However, if Active Control is available then a check is needed for its activation. The activation of Active Control is possible after any number of Ad-Hoc Control procedures according to the following formula:

$$[TAR \times CAR \times (n_i + n_r)] \times k > CACS \tag{4}$$

where k is an activation coefficient.

n_i is the number of affected items in the i^{th} Ad-Hoc attempt.

n_r is the number of remaining items that have to be processed during delay of Active Control.

An additional check has to be performed if $n_r < \frac{PID}{2}$, then Active Control is suspended (called off) even if the condition in Equation (4) holds. PID is a number of items which can be processed during delay of Active Control. For example, if 'Delay of Active Control' is 1 week, PID = 230. Then, if after 1 Ad-Hoc Control procedure, 100 affected items are left, no Active Control is activated.

If the condition in Equation (4) is true, then Active Control will be applied. Otherwise the Ad-Hoc Control procedure is used. In case of applying the Ad-Hoc Control procedure, NC = 0 and no SPOT properties are affected. Since Active Control incurs a delay, the cost of dealing with risk is calculated as follows:

$$cost_{risk} = cost_{ad-hoc} + cost_{ad-hoc-pid} + CACS \tag{5}$$

where $cost_{ad-hoc}$ is the cost of the Ad-Hoc Control procedure before the activation of Active Control.

$cost_{ad-hoc-pid}$ is the cost of the Ad-Hoc Control procedure during delay before Active Control has effect.

Active control will be activated for a given activity only if a sufficiently large number of items will pass through an activity. For the example workflow scenario discussed in the following section, a threshold of 100 files was chosen according to the practices at ORF.

VII. BPRisk Application and Results

In this section, we give an example of how the BPRisk framework has been applied in the design of a workflow in collaboration with the Austrian Broadcasting Corporation, ORF. Respective sections below present a workflow, simulation scenarios, results from simulation modelling and an discussion by ORF to evaluate the accuracy and value of the simulation results.

A. MXF Workflow

Within the DAVID project, the BPRisk framework has been developed with use cases from both the French National Archive (INA) and the Austrian Broadcasting Corporation (ORF), such as planning for migration of old, analogue, content into new, digital, formats (digital migration). Here, we include an example of the use of BPRisk in the planning of an MXF Repair workflow at ORF, which has been used within the DAVID project for validation purposes. MXF is an abbreviation for a file format; Material eXchange Format. The standard for its use is ambiguous in places and some tool implementations are inconsistent. The result is format compatibility issues, i.e., the files may not standard compliant and, therefore, may not be possible to play in the future. The MXF workflow uses a service called CubeWorkflow [33], which analyses media files for compatibility issues at the file wrapper and bit stream levels. For this scenario MPEG-2 [34] encoded (bit stream) MXF (wrapper) D-10 (SMPTE 356M) [35] files were used. A logical view of the different layers of a digital AV file is illustrated in Figure 10. After the workflow design (planning) was completed, the workflow was executed and the results of the planning could be compared with the monitoring data collected during its execution (see Section VII-D, below).

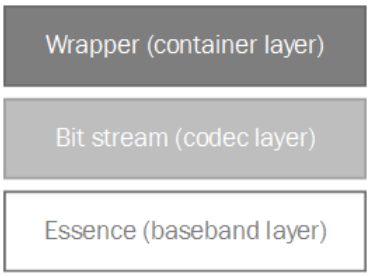


Figure 10. Logical view of layered structure of AV files.

The MXF Repair workflow is depicted in Figure 11, which consists of 9 activities and 2 exclusive OR gateways (both pertaining to points where errors may be detected and handled) Each activity is briefly described as follows:

TSM Retrieve: an activity representing the retrieval (download) of MXF files from a Tivoli Storage Management system (TSM; an LTO tape based IT storage unit).

CubeTec Repair Server Input-Share: network storage on the CubeTec Repair Server to store the retrieved MXF files from the previous activity in order to be processed by the Cube Workflow system.

Cube Workflow: this activity represents a black-box of the Cube Workflow system executing a general analysis (of the Wrapper and Streams) of the MXF files in the INPUT-Share (previous activity). This analysis will detect relevant file errors, attempt to repair errors, and conduct a final control/analysis check of the repaired files. The MXF files passing the final check will be transferred to the ESYS Input-Share.

ESYS Input-Share: network storage in the ESYS Server-framework to gather files for Upload. ESYS = Essence Storage System (by IBM), used by ORF.

Upload: general storage step of ESYS, where MXF files are written to LTO tapes in ESYS and registered in FESAD (the TV Archives MAM), including the production and registration of preview files and keyframe light tables.

Mapping Control: this activity is a manual quality control of the automatic mapping results from the upload process, which is conducted by a person. This is done by comparing the file content (video, audio) with the descriptive metadata in the FESAD entry.

Repair: this is an optional step, if the previous Mapping Control revealed a mapping error; manual "reallocation" of the affected file(s) from the wrong FESAD entry to the correct one.

Preview Alignment: manual setting of correct IN and OUT markers of video footage via a special Preview Alignment Tool to mark the beginning and end of specific sections. This step is needed to avoid wrong preview ranges in clustered contents and alike.

Repair / Adjustments: this final activity covers other necessary manual repairs in descriptive and technical metadata (this is a general activity at ORF whenever major changes are done in a FESAD entry).

This is a small workflow, which is ideal for validation and visualisation to help clarify aspects of the risk specification and role of workflow simulation in the BPRisk framework. However, workflows can be significantly larger and more complex, which is a trend we can expect to continue in the future given the adoption of automated tools for carrying out workflow activities in the media industry. This, in turn, will increase the demand for tools to help with workflow planning and analysis.

In the DAVID project, the DOMAIN layer of the BPRisk ontology has been created based on controlled vocabularies for preservation activities, tools and risks, which has been discussed above in Section V-C. Each of the workflow activities have been mapped to activities in this controlled vocabulary. For example, the first activity in the workflow, 'TSM Retrieve', maps to 'Acquisition/Recording' in the preservation vocabulary. And two risks have been identified for this activity: a) wrong file selection and b) retrieve fails. The semantic reasoning rules discussed above, in Section V, enables the BPRisk framework to prompt users with such risks at design time when they add the activity to the workflow they are designing.

After specifying risks for the different activities, workflow simulation scenarios were set up with ORF for this workflow. To simulate workflow execution, additional parameterisation is required, such as estimates for how often the risks are likely to occur, and the expected time and costs for dealing with any issues that may occur. Values were set based on the experiences the workflow and technical experts at ORF have of the tools and activities used in the workflow, as well as observations from monitoring data where available. In the future, these estimates are intended to be updated and improved via the Risk Feedback Centre, as discussed above in Section IV-C.

B. Risk Simulation Configurations

In the MXF Repair workflow, each activity has got two risks that can occur. The risks are a mixture of system and human errors. For example, the 'Retrieve Fails' risk for the 'TSM Retrieve' activity is caused by technical error, while the

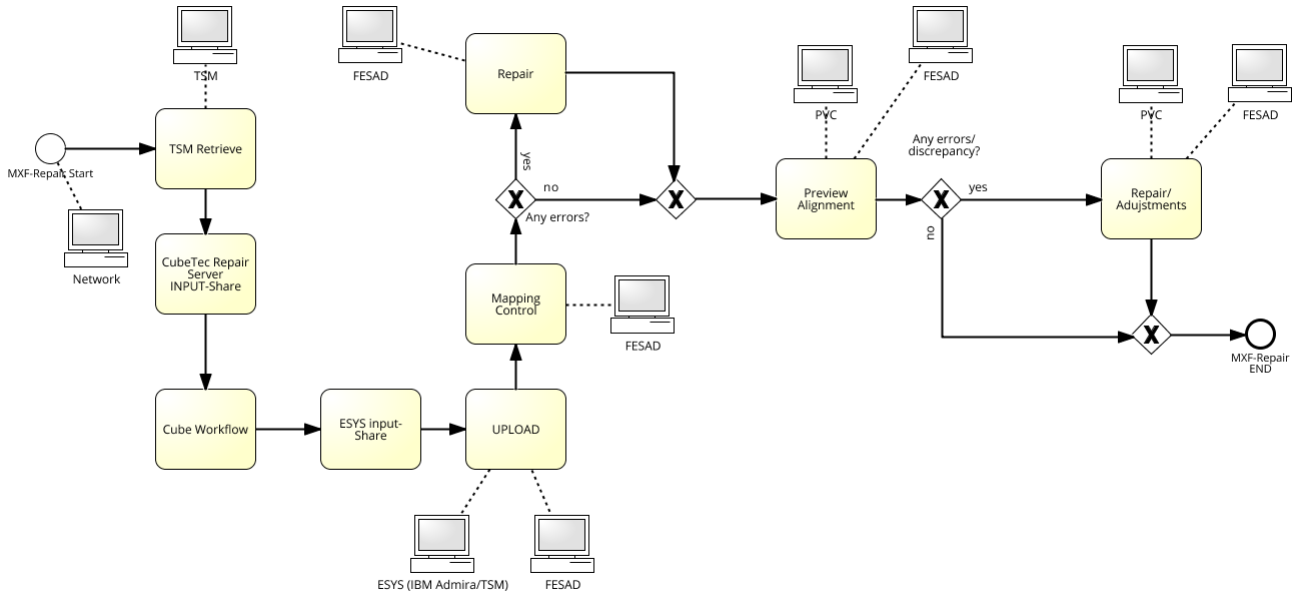


Figure 11. MXF Repair workflow.

'Wrong Assessment' risk for the 'Mapping Control' activity is caused by human error.

For some activities, the two risks are mutually exclusive, i.e., they cannot occur simultaneously; e.g., for the 'TSM Retrieve', 'Cube-Tec Repair Server INPUT-Share' and 'ESYS input-Share' activities. The activity 'Upload' has got two risks that can either happen independently or simultaneously, which requires a combined frequency risks occurrence to be provided for this activity. Moreover, if the risk 'Copy Error' takes place in the previous activity ('ESYS Input-Share'), then the frequency of the risk 'Fails' (for the 'Upload' activity) increases slightly (temporally). For the rest of the activities, risks have the following dependencies: initially the risks are modelled as mutually exclusive, however, if the 'Overload' risk takes place then the frequency of the second risk increases by the given Multiplication Factor and occurrence of this second risk is simulated with the new frequency. That is, if 'Overload' occurred, it can cause the second risk such as wrong assessment or wrong mapping to take place too.

After the activity 'Mapping Control', the simulation procedure will check whether any errors took place (the first XOR gate). If there are errors, then the 'Repair' activity is performed. The probability of any errors is 0.1%, i.e., the 'Repair' activity will take place very seldom during the workflow runs. A condition after the 'Preview Alignment' activity checks whether any error or discrepancy is present. It is known that probability of this error/discrepancy is 60%.

The Estimated Frequencies for each risk are estimated based on an annual throughput of approximately 12,000 files (1,000 monthly, 230 weekly, 46 daily). It is assumed that the MXF Repair workflow runs 15 working hour per day, 5 days a week, 22 working days in a month and 264 working days in a year.

C. Simulation Results

Two simulation scenarios were explored for this workflow, since ORF were interested in a comparison between what is currently done to control risk and a worst-case situation in which no risks were controlled. The 'no control' scenario in this sense serves to demonstrate the importance of what is currently done. For this purpose we ran simulations with 1,000 items (files) processed via a workflow 10,000 times. Given the throughput figures presented above, this equates to 22 days of executing the workflow. The results discussed here are focused on demonstrating the type of insights the simulation modelling gives users in terms of performing risk management for business process workflows. Due to commercial sensitivity, certain parameters such as the financial cost of dealing with risk, are not accurate.

When a simulation is executed in the BPRisk framework, the UI has been designed to give an overview of key facts. For example, for a simulation scenario of the existing set-up for controlling risk:

Risk affecting the most items (**88**): **Upload Fails**.

Total cost of risk treatments: **€1,424.93**.

The most expensive risk (**€965.16**): **Upload Fails**.

Total time spent on dealing with risk: **26.13 hours**.

The most time consuming risk: **Upload Fails**.

Main issue caused is loss of **Authenticity**.

In terms of identifying key vulnerabilities and determining where to target investments, the above summary gives a strong indication towards the 'Upload' activity. From that point, users can explore the data in more detail. Figure 12 shows how many times risks occurred for each activity and how many risks were undetected overall (mean values for the 10,000 Monte Carlo runs). Although the 'Fail' risk for the 'Upload' activity was flagged in the summary as the most expensive risk, both

in terms of time and cost, this figure shows that 'Preview Alignment' had the most risk occurrences (followed by 'Upload' and 'Mapping Control'). 'TSM Retrieve' and 'Cube-Tec Repair Server Input-Share' have the smallest number of risk occurrences. No risks took place for the 'Repair' activity, since it is a rare event and it was not evoked even a single time during the simulation. We also observe that a certain proportion of risks are be undetected. This will incur SPOT impacts, which we will return to below.

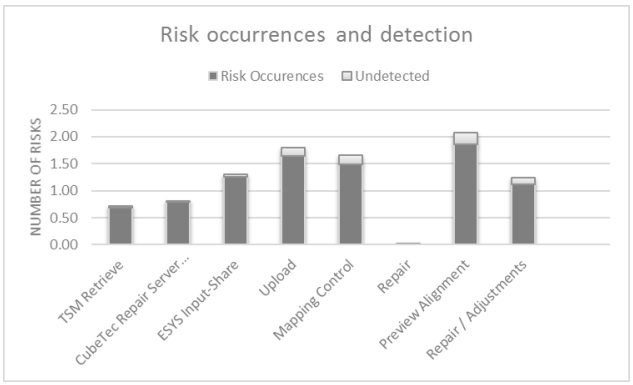


Figure 12. Risk occurrences and detection.

Although the 'Preview Alignment' activity incurs the most risk instances, the risks occurring during the 'Upload' activity will affect the largest number of items; 88 files on average (out of the 1,000). This is significantly more than any other activities, which is depicted in Figure 13. In comparison, the number of affected items during 'Preview Alignment' was 6 on average, and 3 for 'Mapping Control'.

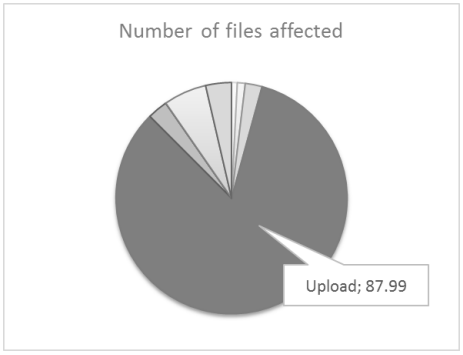


Figure 13. Number of files affected by risk.

Next, we can investigate how the Ad-Hoc Control procedure and Active Control cope with these risks. Figure 14 shows a sum of Negative Consequences (NCs) for each activity with and without Ad-Hoc Control. NC is calculated based on the 'Level of Severity' parameter (defined in the range {1,3}), by summing up this pre-defined value for each risk occurrence (for each activity) that was undetected. For example, the LS for the Upload Fails risk is 1 and the risk occurred 1 times on average, giving a NC value of 1 without control. Mapping Control risks have defined LS as 2, and the 'Wrong Assessment' risk occurs 0.63 times on average, giving a NC value of 1.26 without control.

Not surprisingly, the difference in NC between the two scenarios is significant. If no Ad-Hoc Control procedure is in a place, then the largest NC from risks will be for 'Mapping Control', 'Preview Alignment' and 'ESYS Input-Share'. The Ad-Hoc Control procedure reduces NC to zero for all risks that have been detected. However, for risks such as 'Wrong Assessment' for the 'Mapping Control' activity, the Detection Level is set to 75%, giving a small NC value of 0.32 as the risk only occurred 0.63 times on average during the simulation. The top three activities in terms of the highest NC is almost identical when Ad-Hoc Control is applied; the difference is that the 'Upload' activity has replaced 'ESYS Input-Share'.

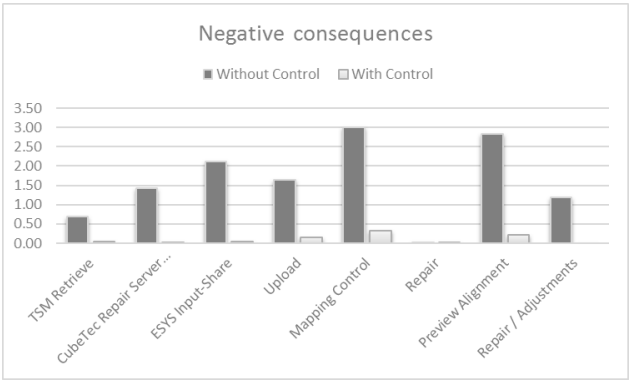


Figure 14. Negative consequences with and without Ad-Hoc Control.

Undetected, uncontrolled, risks will have an impact, which we represent according to the SPOT model (discussed above in Section VI-B). For this simulation, the SPOT impact can be seen in Figure 15, comparing the two scenarios with– and without control. The values in this figure are calculated similarly to NC, above, by summarising the number of times risks have been either undetected or have not been successfully "fixed" in a control procedure. This gives an appreciation of the types of issues to the digital content, caused by the various risks. For example, when risks are controlled, the main issue is loss of authenticity, which is interesting for this type of workflow. Authenticity, as described in Section VI-B refers to the digital content indeed being what it purports to be. Considering that this workflow addresses format compatibility issues, which includes issues such as incorrect or inconsistent meta-data to describe digital content, authenticity being the main issue despite controlling the known risks is interesting as it reflects that the repair processes itself is not 100%.

The Ad-Hoc Control procedure is successful for this MXF Repair workflow in dealing with risks. As seen in Figure 15, the SPOT impacts are minor when the risks are controlled. However, controlling risks incurs costs, both in terms of time to deal with the risk occurrence, as well as financial cost. Figure 16 shows a pie chart that illustrates well the proportional differences in financial cost for addressing risks for the different activities, in which the Upload activity accounts for approximately 68% of the total costs. Moreover, addressing the risks to this activity takes 19.30 hours on average, which is approximately 2.5 working days out of the 22 days of simulating this workflow execution.

Active Control was only activated once, during the 'Repair Adjustment' activity. This was for the risk 'Overload', which

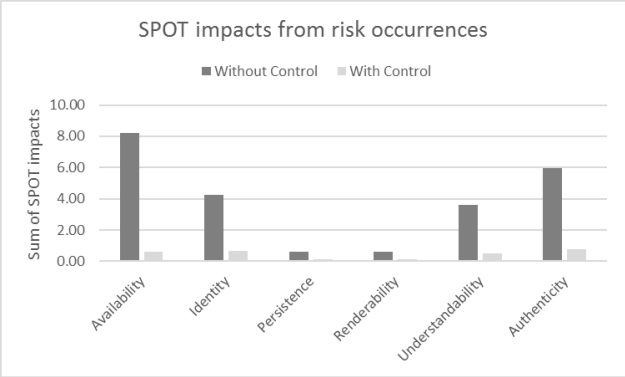


Figure 15. SPOT properties affected by risk (impact - with and without control).

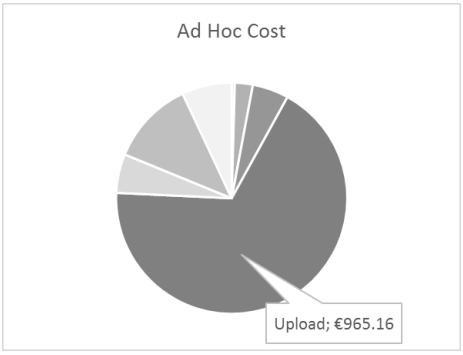


Figure 16. Cost of Ad-Hoc Control.

resulted in a cost of €500 in this simulation scenario. Although the Ad-Hoc Control procedure of dealing with risks is very effective in this particular workflow, it is of interest to rank risks according to impact on a workflow based on a generic score. Table II shows the ranking of risks according to impact score, which was calculated as the product of the number affected items and negative consequences, based on the mean values from the 10,000 simulation runs of the ORF MXF Repair workflow under condition that no Ad-Hoc Control procedure was in place. The ranking is an approach to presenting the risk simulation results in a way to help identify prioritisations of which risks to address. Not only does this representation identify which activities are the most vulnerable to risk, but also we can see that the top 5 highest ranked risks are of the same type - overload. This kind of information, therefore, indicates there is a common technological issue with the systems used not being able to handle the workflow. There may, instead, be alternative systems available, which could be considered in different workflow scenarios to then compare potential ROI, e.g., in terms of whether there is financial gain in upgrading a system if it reduces the risk occurrences.

D. Evaluation and Discussion of Simulation Results

Based on the results discussed above, this section is an evaluation technical experts (workflow designers) from ORF based on the observations they have made in reality. Both ranking of risk/activities and the impact is giving an exact picture of the actual situation ORF experienced when running the MXF Repair workflow. For example, they experienced several upload-fails and 'Overload' was indeed the greatest issue in 'Preview-Alignment'.

Even the ranking from the simulation results (Table II) is nearly identical to ORF's experience. The top three is identical. Thereafter, there are just a few instances of risks that have a rank difference of 1 place. Even at the very end of the ranking the results match ORF's experience, which impressed their workflow-designers very much. Also, in this section of the results, there is only one swap in ranking positions. In the actual workflow 'Retrieve Fails' was slightly higher than 'Wrong File selected'. This may be due to the fact that in the early part of the MXF Repair process they experienced some severe network-issues.

The results for risk occurrence and the number of affected items show a similar picture; the results of the simulation match with our experience during the actual workflow execution. Just the absolute number of affected items for 'Upload' 'Fails' seems to be too high; a first explanation may be the fact that in the actual workflow those upload failures had a significant concentration in the first 6 months of the MXF Repairs (due to several changes in the affected ESYS-systems) – and from the same time frame the values for the simulation-model has been taken.

The effect of Ad-Hoc Control, shown in Figure 14, is a very strong argument and help for implementing a proper instalment for control measurements. And again the figures for NC in the different activities reflect very well the actual situation in the MXF Repair workflow; very interesting is the rather high effect of "technical" measurements (e.g., in ESYS Input Share) compared to those with "human-related" measurements (e.g., Mapping Control).

The high impact of Ad-Hoc Control reflects again the actual experience and strengthens our arguments in "investing" in those by having a rather large overhead budget available. Especially the results in the SPOT model for Ad-Hoc Control are very impressive and promising. Finally, the results for costs are significant assets for putting forth arguments in budget discussions for future workflows in the domain. The extremely positive impact of Ad-Hoc Control was always neglected or doubted by the decision-makers at ORF.

It has to be stressed here, that the quality of the results are based not only on the model, but also on the quality of the input-data given. ORF put significant effort and time in providing accurate and reliable data for both the model and the simulation, so it has to be expected that the old archive-rule "Garbage in = Garbage out" is valid for risk management and assessment as well. You have to invest a good percentage of the budget reserved for "accompanying measures" for this activity (in case of the MXF Repair workflow, approximately $\frac{1}{3}$), which included the calculation and surveying of quality planning data to get good results and actually have the chance to save money in the actual workflow or process. And being an expert or involving experts in the domain is highly necessary to save efforts in this process of planning and data collection.

For all colleagues at ORF involved in the process, the tool proves to be a very good and highly reliable instrument to evaluate risks and their actual impact even before or at an early stage of the implementation of a workflow. However, as discussed above, if the model is not configured with the right

TABLE II. Risk ranking according to simulation results and ORF - in descending order.

Activity	Risk	Impact	Score	Rank	ORF Rank
Upload	Fails	Very high	86.99	1	1
Preview Alignment	Overload	High	8.58	2	2
ESYS Input-Share	Overload	High	3.46	3	3
Mapping Control	Overload	High	3.46	4	5
Repair / Adjustments	Overload	High	3.09	4	4
Preview Alignment	Wrong assessment	Medium	1.75	5	7
Cube-Tec Repair Server Input-Share	Overload	Medium	1.28	6	6
Mapping Control	Wrong assessment	Medium	1.28	7	7
Upload	Wrong parameters	Medium	0.64	8	8
ESYS Input-Share	Copy error	Low	0.19	9	9
TSM Retrieve	Wrong file selected	Low	0.14	10	11
TSM Retrieve	Retrieve fails	Low	0.14	11	10
Cube-Tec Repair Server Input-Share	Copy error	Low	0.03	12	12
Repair / Adjustments	Wrong assessment	Very low	0	13	13
Repair	Wrong mapping	Very low	0	13	14
Repair	Overload	Very low	0	13	14

(and accurate) data, the results will not match with reality. Therefore, it is important to complete the risk management process, based on the Deming cycle, to capture monitoring information from the workflow executions to update both the risk information and simulation configuration to improve the accuracy.

VIII. Conclusion and Future Work

We have presented a Business Process Risk management framework (BPRisk) that allows users to manage workflow processes with regards to risk in order to reduce cost and increase the long-term value of digital media content. The framework is generic in nature, but has been discussed here in the context of digital preservation, where the objective is to avoid damage to the digital content and ensuring that the content can be accessed in the future.

The BPRisk framework combines workflow specification and risk management. It has been designed in accordance with a risk management process based on the Deming (PDCA) cycle and we have shown how it relates to the stages of the ISO 31000 risk methodology. Long-term digital preservation is threatened by format obsolescence, media degradation, and failures in the very people, processes and systems designed to keep the content safe and accessible. Therefore, investing in substantial planning and design is essential in order to prevent issues that may not be possible to rectify; rendering the content unusable. Further, key to the risk management process is continual improvement, i.e., risk management is not merely a static exercise performed at design time [13], but it is also imperative during process change.

A layered semantic risk model has been presented, which a) enables reasoning about threats in a workflow and b) assists end-users (who are typically not risk experts) by automatically suggesting relevant risks and respective controls for workflow activities. This helps the workflow designers specify more robust workflows, reducing the risk of causing irretrievable damage to the media content. Moreover, the framework helps workflow designers optimise workflows and improve cost-benefit by identifying (and addressing) key vulnerabilities by simulating workflow executions to estimate the impact of risk.

The simulation service in the BPRisk framework allows users to estimate the impact of risks before executing the workflow, which increases the chances of detecting issues rather than jeopardising the real media content. A workflow simulation provides users with information about *inter alia* the quantity of media content that may be affected by risk (and how), and the time and cost of dealing with risk (i.e., control and treatment). Therefore, organisational impacts can be derived and users may simulate different what-if scenarios in order to evaluate different workflow designs before proceeding with executing a particular workflow process on live material. What-if scenarios may, e.g., involve justifying reasons for putting in place certain control mechanisms by showing that the ROI is positive.

A prototype of the BPRisk framework has been developed in the DAVID project. To demonstrate the application of the framework and the value of the simulation results, we have reported on an evaluation scenario with the Austrian Broadcasting Corporation (ORF). The technical experts at ORF found the results to be almost identical to what they have observed by executing the workflow. Key benefits emphasised include: a) investing time in workflow planning and controlling risks in order to prevent issues, and b) justifying workflow designs and risk controls to decision makers.

Further research involves implementing mechanisms for automatically updating risk models and respective simulation configurations according to observed workflow execution data in order to improve the support for continual improvement of workflow processes.

Acknowledgements

This work has been carried out in the DAVID project, supported by the EC 7th Framework Programme (FP7-600827).

References

[1] V. Engen, G. Veres, S. Crowle, M. Bashevoy, P. Walland, and M. Hall-May, "A Semantic Risk Management Framework for Digital Audio-Visual Media Preservation," in The 10th International Conference on Internet and Web Applications and Services (ICIW). IARIA, June 2015, pp. 81–87.

[2] EC FP7 DAVID Project, "Digital AV Media Damage Prevention and Repair," http://david-preservation.eu/, [retrieved: 2016.02.29].

[3] Object Management Group, "Business Process Model and Notation (BPMN) Version 2.0," http://www.omg.org/spec/BPMN/2.0/PDF/, [retrieved: 2016.02.29].

[4] Department of Special Collections, Donald C. Davidson Library, University of California, "Cylinder Preservation and Digitization Project," http://cylinders.library.ucsb.edu/, [retrieved: 2016.02.29].

[5] V. Engen, G. Veres, M. Hall-May, J.-H. Chenot, C. Bauer, W. Bailer, M. Höffernig, and J. Houpert, "Final IT Strategies & Risk Framework," EC FP7 DAVID Project, Tech. Rep. D3.3, 2014, available online http://david-preservation.eu/wp-content/uploads/2013/01/DAVID-D3.3-Final-IT-Strategies-Risk-Framework.pdf [retrieved: 2016.02.29].

[6] M. Addis, R. Wright, and R. Weerakkody, "Digital Preservation Strategies: The Cost of Risk of Loss," SMPTE Motion Imaging Journal, vol. 120, no. 1, 2011, pp. 16–23.

[7] J.-H. Chenot and C. Bauer, "Data damage and its consequences on usability," EC FP7 DAVID Project, Tech. Rep. D2.1, 2013, available online http://david-preservation.eu/wp-content/uploads/2013/10/DAVID-D2-1-INA-WP2-DamageAssessment_v1-20.pdf [retrieved: 2016.02.29].

[8] D. Rosenthal, "Format Obsolescence: Assessing the Threat and the Defenses," Library Hi Tech, vol. 28, no. 2, 2010, pp. 195–210.

[9] M. Addis, "8K Traffic Jam Ahead," PrestoCentre Blog, April 2013, available online: https://www.prestocentre.org/blog/8k-traffic-jam-ahead [retrieved: 2016.02.29].

[10] ISO/IEC, 31000:2009 Risk management - Principles and guidelines, ISO Std., 2009.

[11] D. Green, K. Albrecht, and et al, "The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials," National Initiative for a Networked Cultural Heritage, Tech. Rep., 2003, available online: http://www.ninch.org/guide.pdf [retrieved: 2016.02.29].

[12] M. Hall-May, B. Arab-Zavar, J. Houpert, C. Tiensch, H. Fassold, and V. Engen, "Analysis of loss modes in preservation systems," EC FP7 DAVID Project, Tech. Rep. D2.2, 2014, available online http://david-preservation.eu/wp-content/uploads/2013/01/DAVID-D2.2-Analysis-of-Loss-Modes-in-Preservation-Systems.pdf [retrieved: 2016.02.29].

[13] S. Suriadi, B. Weiß, A. Winkelmann, A. Hofstede, M. Adams, R. Conforti, C. Fidge, M. La Rosa, C. Ouyang, A. Pika, M. Rosemann, and M. Wynn, "Current Research in Risk-Aware Business Process Management - Overview, Comparison, and Gap Analysis," BPM Center, Tech. Rep. BPM-12-13, 2012.

[14] M. Rosemann and M. zur Muehlen, "Integrating Risks in Business Process Models," in ACIS Proceedings, 2005.

[15] A. Sienou, E. Lamine, A. Karduck, and H. Pingaud, "Conceptual Model of Risk: Towards a Risk Modelling Language," in Web Information Systems Engineering, ser. LNCS 4832, 2007, pp. 118–129.

[16] R. Conforti, G. Fortino, and A. t. M. La Rosa, "History-Aware, Real-Time Risk Detection in Business Processes," in On the Move to Meaningful Internet Systems, ser. LNCS. Springer, 2011, vol. 7044, pp. 100–118.

[17] X. Bai, R. Krishnan, R. Padman, and H. Wang, "On Risk Management with Information Flows in Business Processes," Information Systems Research, vol. 24, no. 3, 2013, pp. 731–749.

[18] M. Addis, M. Jacyno, M. H. Hall-May, and S. Phillips, "Tools for Quantitative Comparison of Preservation Strategies," EC FP7 PrestoPRIME Project, Tech. Rep. D2.1.4, 2012, available online: http://eprints.soton.ac.uk/349290/ [retrieved: 2016.02.29].

[19] V. Lloyd, ITIL Continual Service Improvement – 2011 Edition. The Stationary Office, 2011, iSBN: 9780113313082.

[20] Signavio GmbH, "Signavio Decision Manager," http://www.signavio.com/products/decision-manager/, [retrieved: 2016.02.29].

[21] JBoss, "jBPM," http://www.jboss.org/jbpm, [retrieved: 2016.02.29].

[22] L. Richardson and S. Ruby, RESTful Web Services, 1st ed. O'Reilly, May 2007.

[23] Pivotal Software, "Spring," https://spring.io/, [retrieved: 2016.02.29].

[24] G. E. Krasner and S. T. Pope, "A cookbook for using the model-view controller user interface paradigm in Smalltalk-80," Journal of Object-Oriented Programming, vol. 1, no. 3, Aug/Sept 1988, pp. 26–49.

[25] Activiti, "Activiti BPM Platform," http://activiti.org/, [retrieved: 2016.02.29].

[26] Mathworks, "Simulink," http://uk.mathworks.com/products/simulink/, [retrieved: 2016.02.29].

[27] L. Moreau and P. Missier, "PROV-DM: The PROV Data Model," W3C Recommendation, 2013, available online: http://www.w3.org/TR/2013/REC-prov-dm-20130430/ [retrieved: 2016.02.29].

[28] M. Surridge, A. Chakravarthy, M. Hall-May, X. Chen, B. Nasser, and R. Nossal, "SERSCIS: Semantic Modelling of Dynamic, Multi-Stakeholder Systems," in 2nd SESAR Innovations Days, 2012.

[29] F. Bellard, "FFmpeg," https://www.ffmpeg.org/, [retrieved: 2016.02.29].

[30] W3C, "RDF 1.1 Turtle," W3C Recommendation, 2014, available online: https://www.w3.org/TR/turtle [retrieved: 2016.02.29].

[31] ——, "SPARQL Inferencing Notation (SPIN)," W3C Submission, 2011, available online: http://www.w3.org/Submission/2011/02/ [retrieved: 2016.02.29].

[32] S. Vermaaten, B. Lavoie, and P. Caplan, "Identifying Threats to Successful Digital Preservation: the SPOT Model for Risk Assessment," D-Lib Magazine, vol. 18, no. 9/10, 2012.

[33] Cube-Tec International, "Cubeworkflow," https://www.cube-tec.com/en-uk/products/workflow/cube-workflow/cube-workflow-20, 2016, [retrieved: 2016.02.29].

[34] ISO/IEC, ISO/IEC 13818-1:2000 Information technology – Generic coding of moving pictures and associated audio information: Systems, ISO Std., Dec 2000.

[35] SMTPE, "For Television – Type D-10 Stream Specifications – MPEG-2 4:2:2P @ ML for 525/60 and 625/50," SMPTE ST 356:2001, Aug 2001, pp. 1–7.