# A Semantic Risk Management Framework for Digital Audio-Visual Media Preservation

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

***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. In this paper, we focus on dealing with issues resulting from system errors, rather than random failure or corruption; issues caused by the people, systems and processes that handle the digital AV content. We present a semantic risk management framework designed to support preservation experts in managing workflow risks, combining workflow and risk specification within a risk management process designed to support continual improvement of workflow processes.**

***Keywords–Risk management; semantic modelling; business processes; workflows; media preservation; digital archives.***

## 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 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 contents 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.

A business process risk management framework (BPRisk) has been developed in the EC FP7 DAVID project [1], 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, but focus on the semantic modelling and risk specification aspects. Within the DAVID project, this research and development has been conducted to provide a tool to help prevent damage to digital AV content. In addition to this, the DAVID project focuses on understanding damage (how it occurs and its impact), detecting and repairing damage, and improving the quality of digital AV content.

The BPRisk framework is generic in nature, supporting risk specification for Business Process Modelling Notation (BPMN) 2.0 [2] 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 doing risk analysis, but semantic reasoning is utilised to enable suggestions on risks and controls for the activities in the workflows at design time.

In the remainder of this paper, we will further discuss the challenges and related work on digital preservation in Section II and risk management in this domain in Section III. Thereafter, in Section IV, we present the BPRisk framework, followed by details of the semantic modelling adopted in the framework in Section V. Section VI discusses the implementation status of BPRisk and a real application from within the DAVID project. Section VII concludes this paper and discusses further 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. Some content needs to be retained and archived to

 ISBN: 978-1-61208-412-1

enable content re-use, e.g., for cultural heritage, or due to regulatory compliance for security, health and safety. 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 [3]. 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) in the DAVID project undergo digital migration projects to digitise the older, analogue, content [4]. 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."* [4]

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 [5], [6], [7].

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 [5], which can effectively be addressed by keeping multiple copies of each file [4], [6]. We also need to consider the future challenges for digital preservation as some analyses [8] predict that as ever more 8K AV content is ingested into archives, the growth in data volumes will, with all likelihood, outstrip the growth in storage capacity and increase in data write rate, such that it becomes impossible to store and replicate all 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 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"* [9]. Risk, as defined by ISO 31000 [9], 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 broadcaster ORF typically deploy a number of IT based strategies for avoiding, preventing or recovering from loss [4]. These archives are engaged in a process of long-term Digital Asset Management (DAM) [10], 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.

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. [11], 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) [12]. 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. [13]

 ISBN: 978-1-61208-412-1 

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. [14] 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. [15] 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 [16]. 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 Return On Investment (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 of change 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 [9], 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 [17]. In general terms, risk management is a part of continual improvement of processes – preservation workflows in this context.

The ISO 31000 [9] 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

 ISBN: 978-1-61208-412-1 

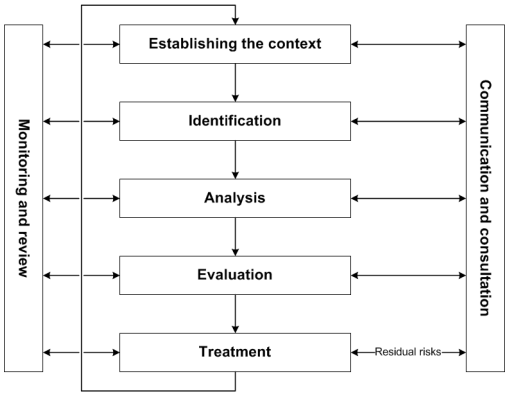


Figure 1. ISO 31000 risk management process.

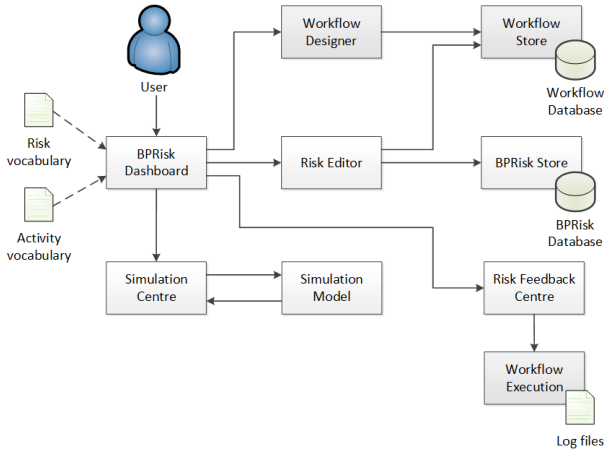


Figure 2. BPRisk framework high level component view.

is obsolescence and one of the key current risk strategies involving file-replication may not be feasible in the future.

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. [11]; 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 is implemented as a RESTful web application, integrating 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 [18] and the jBPM Designer [19]. 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.

## V. Semantic Risk Modelling

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

 ISBN: 978-1-61208-412-1

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), can be extended in unexpected ways, and it can be combined with other ontologies such as W3C PROV [20].

The approach to building the ontology is based on work done in SERSCIS project [21]. 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 subclasses 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 who may not have this expert knowledge themselves when they are designing new systems.

The same, layered, ontological approach has been taken here, but the core ontology is slightly different. While, in SERSCIS, the triad in the CORE layer 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.

### B. Model Definition

The model focuses on the Activities in the preservation lifecycle 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 SPIN rules (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 3.

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 3 include two sub-classed Activities, 'Migration' and 'Digital Migration', with an associated risk 'Migration Fails'.

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 Risk to them. 'FFmpeg Migration' is given as an example in Figure 3

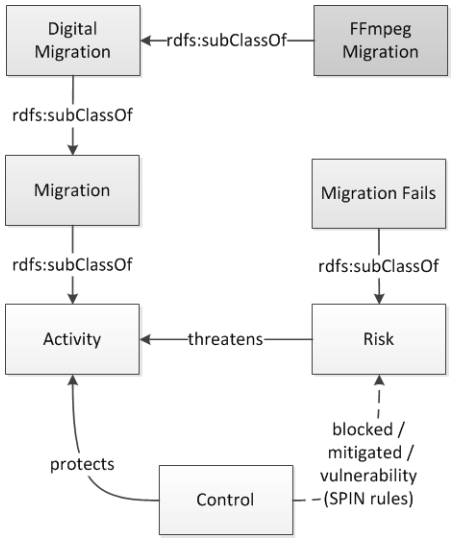


Figure 3. BPRisk ontology with sub-classing examples.

(dark grey), which is a subclass of 'Digital Migration'. This digital migration risk is specific to using the tool FFmpeg [22], which is a common AV media transcoding tool. This sub-classing is important, as we can reason about risks throughout the hierarchy, which we go further into below.

### C. Semantic Reasoning

The relationships between risks, controls and activities are encoded as risk classification rules using SPIN [23]. Running inferencing over the model automatically does the classification. For example, the following SPIN rule classifies an instance of the risk *FieldOrderIssues*, which threatens the activity *Transcoding*, as blocked, if the control *ChangeTranscodingTool* is present:

```
CONSTRUCT {
    ?r a dom:BlockedRisk .
} WHERE {
    ?a a act:Transcoding .
    ?r a dom:FieldOrderIssues .
    ?c a dom:ChangeTranscodingTool .
    ?r core:threatens ?a .
    ?c core:protects ?a .
}
```

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 3. This should specify the kind of activity in the preservation workflow of interest, e.g., subclass *Scanning* as *35mmToJPEG2kScanning*. Workflow-specific risks can then be automatically generated using SPIN. For example, the following is a generic SPIN 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 .
```

 ISBN: 978-1-61208-412-1 

```
        ?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.

Encapsulation of media preservation knowledge (linking activities, risks and controls) using SPIN rules provides a flexible and extensible representation of knowledge based reasoning in our architecture. Specifically, we extend the SPIN templates rule hierarchy into which we insert groups that can contain rules to be called upon, for example, in the construction of new risk instances in the presence of particular activities. Using this approach, it is possible to progressively refine the core preservation knowledge base (or augment it with additional, domain specific rules) without necessarily updating system code with new SPARQL [24] queries.

## VI. BPRisk Implementation and Application

At the time of writing, the BPRisk framework prototype has been developed within the DAVID project [1]. It has been implemented as RESTful web service using Java Spring [25]. As noted in Section IV-C, above, the jBPM Designer [19] has been integrated for workflow design. A risk simulation model has been implemented in Matlab Simulink [26]. The Risk Feedback Centre is under development.

Within the DAVID project, the BPRisk framework has been developed with use cases from INA and 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 are not standard compliant and may not be possible to play in the future. After the workflow design (planning) was completed, the workflow has been executed and the results of the planning could be compared with the monitoring data collected during its execution.

The MXF Repair workflow is depicted below in Figure 4. Due to space restrictions, a description of the workflow activities is not included here, nor are some application/modelling details. The aim here is to clarify aspects of the workflow risk specification and the role of workflow simulation. However, interested readers are referred to specific parts of [4], below, for further details.

Firstly, in the DAVID project, the DOMAIN layer of the BPRisk ontology has been created based on controlled vocabularies for preservation activities, tools and risks. Interested readers can refer to Annex C in [4] for details. The first activity in the workflow, TSM (Tivoli Storage Manager, a data backup system from IBM) Retrieve, maps to 'Acquisition/Recording' in the preservation vocabulary, for example. 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.

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.

The simulation results on this workflow showed very clearly that the activity most affected by risk is the Upload activity (upload fails). Not just in terms of frequency of occurrence, but it affects the most media files and accrues the most significant financial cost. At this stage of the planning phase for designing new workflows, it is such observations that are important in terms of highlighting the key vulnerabilities and start to quantify their impact and potential cost savings by addressing the problems differently. As discussed in Section IV-B, different versions of a workflow may be designed and simulated as part of the planning before making a final, informed, decision and moving to executing the workflow in the real environment. Interested readers are referred to Section 8.4 and 9 in [4] for details on the simulation modelling, results and validation of these results based on the observations made after executing the workflow.

## VII. Conclusions and Further Work

We have presented a Business Process Risk management framework (BPRisk) that allows users to manage workflow processes with regards to risk. 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. 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.

The BPRisk framework combines workflow specification (and adaption) and risk management. It has been designed in accordance to a risk management process presented in this paper, based on the Deming (PDCA) cycle and we have shown how it relates to the stages of the ISO 31000 risk methodology. Key to the process is continual improvement, as risk management is not only a static exercise performed at design time [11], but 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 risks and respective controls for workflow activities. The framework helps end-users develop and optimise workflows, and improve cost-benefit by identifying (and address-

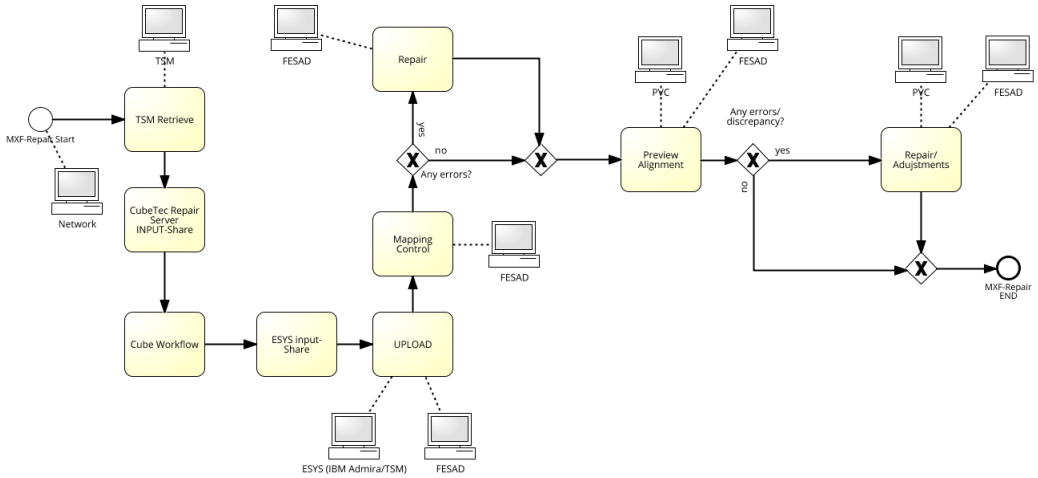


Figure 4. MXF Repair workflow.

ing) key vulnerabilities by simulation workflow executions to estimate the impact of risk.

A prototype is developed in the DAVID project at the time of writing. Further research involves mechanisms for automatically updating risk models and respective simulation configurations according to observed workflow execution data.

## ACKNOWLEDGEMENTS

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

## REFERENCES

[1] EC FP7 DAVID Project, "Digital AV Media Damage Prevention and Repair," http://david-preservation.eu/, [retrieved: May 2015].

[2] Object Management Group, "Business Process Model and Notation (BPMN) Version 2.0," http://www.omg.org/spec/BPMN/2.0/PDF/, [retrieved: May 2015].

[3] Department of Special Collections, Donald C. Davidson Library, University of California, "Cylinder Preservation and Digitization Project," http://cylinders.library.ucsb.edu/, [retrieved: May 2015].

[4] 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: May 2015].

[5] 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.

[6] 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.

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

[8] M. Addis, "8K Traffic Jam Ahead," PrestoCentre Blog, April 2013, available online: https://www.prestocentre.org/blog/8k-traffic-jam-ahead [retrieved: May 2015].

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

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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/.

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

[18] Signavio GmbH, "Signavio Decision Manager," http://www.signavio.com/products/decision-manager/, [retrieved: May 2015].

[19] JBoss, "jBPM," http://www.jboss.org/jbpm, [retrieved: May 2015].

[20] 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: May 2015].

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

[22] F. Bellard, "FFmpeg," https://www.ffmpeg.org/, [retrieved: May 2015].

[23] W3C, "SPARQL Inferencing Notation (SPIN)," W3C Submission, 2011, available online: http://www.w3.org/Submission/2011/02/ [retrieved: May 2015].

[24] ——, "SPARQL 1.1," W3C Recommendation, 2013, available online: http://www.w3.org/TR/sparql11-overview/ [retrieved: May 2015].

[25] Pivotal Software, "Spring," https://spring.io/, [retrieved: May 2015].

[26] Mathworks, "Simulink," http://uk.mathworks.com/products/simulink/, [retrieved: May 2015].

 ISBN: 978-1-61208-412-1 