Towards a High Reliable Enforcement of Safety Regulations - A Workflow Meta Data Model and Probabilistic Failure Management Approach

Heiko Thimm

Abstract—**Today's companies are able to automate the enforcement of Environmental, Health and Safety (EH&S) duties through the use of workflow management technology. This approach requires to specify activities that are combined to workflow models for EH&S enforcement duties. In order to meet given safety regulations these activities are to be completed correctly and within given deadlines. Otherwise, activity failures emerge which may lead to breaches against safety regulations. A novel domain-specific workflow meta data model is proposed. The model enables a system to detect and predict activity failures through the use of data about the company, failure statistics, and activity proxies. Since the detection and prediction methods are based on the evaluation of constraints specified on EH&S regulations, a system approach is proposed that builds on the integration of a Workflow Management System (WMS) with an EH&S Compliance Information System. Main principles of the failure detection and prediction are described. For EH&S managers the system shall provide insights into the current failure situation. This can help to prevent and mitigate critical situations such as safety enforcement measures that are behind their deadlines. As a result a more reliable enforcement of safety regulations can be achieved.**

Keywords— Environmental Health and Safety, workflow management, workflows, failure detection, failure prediction;

I. Introduction

Multiple legal authorities with different responsibility levels obligate companies to follow environmental, health, and safety (EH&S) regulations [1] [2]. Due to the enormous size of this ever growing and frequently revised set of EH&S regulations companies are required to establish an efficient and an effective practice for the enforcement of new regulations and of revisions of existing regulations [3]. Ideally, this enforcement duty is performed trough carefully selected measures such as employee instruction and training, the use of additional safety devices and facilities, and even product revisions in order to reduce the potential of harm [3].

Although there exists a great awareness about the need for a reliable and effective EH&S enforcement practice, often in companies deficiencies can be found in this area [4] [5]. The organizational deficiencies and inappropriate use of Information and Communication Technology (ICT) can create substantial EH&S risks and losses in efficiency and effectiveness. That the use of a Workflow Management System (WMS) [6] will lead to a reliable, effective, and efficient EH&S enforcement in companies seems to be a promising approach.

Traditional WMS are designed to enact and manage the execution of workflow instances according to given workflow models. Typically, the system notifies participants about assigned activities and provides access to information artifacts. However, traditional WMS are not designed to cope with problems that can occur in the context of the enforcement of EH&S regulations. This can lead to an unreliable enforcement of these regulations (i.e. non-compliance) because of activity failures within the execution of EH&S enforcement workflows. As a result critical situations can happen in which the company is threatened by financial losses, by health risks for humans, and by risks for damages to the environment. Activity failures can emerge due to a variety of different reasons. First of all, activity failures can be men made. For example, individuals who are expected to complete EH&S activities can be over-challenged by what is demanded from them. They can also be insufficiently experienced/qualified or suffer from human factors. Activity failures can also be caused by problems inherent to group work such as a bad group atmosphere and group thinks effects. Malfunctioning and defects of components of the corporate technical infrastructure also have to be considered as potential sources for activity failures.

The reliability of EH&S workflow completions can significantly be improved through the use of a WMS that is capable to detect in realtime activity failures that occurred already or that are likely to occur (i.e. prediction) in the near future. Such an enhanced system approach can help companies to prevent and/or mitigate the potential harm resulting from the failures. This consideration presents the overall research objective of the project described in this article. The research is part of a broader project that investigates the integrated use of both WMS and EH&S Compliance Information Systems in order to improve reliability of EH&S regulation enforcement. The focus of the article is on the foundation of this integrated approach which is a domain-specific workflow management meta data model. The modelling concepts of this model are specialized to the detection and prediction of activity failures. The modeling concepts are directed at the company specific organizational context, failure statistics, and proxy templates for real world activities. In the article the modelling concepts are exemplified through a concrete workflow example. Furthermore, an overview of prototypical implementation is presented.

The remainder is organized as follows. An overview of related work is contained in Section 2. The domain-specific meta data model is described in Section 3. Examples of some major modeling concepts of the model are presented in Section 4.

Section 5 describes the main principles of the detection and prediction of activity failures and Section 6 gives an overview of a prototypical implementation. Concluding remarks are contained Section 7.

II. Related Work

In the literature several projects are described that target the monitoring of workflows in order to detect compliance violations [7] [8]. An overview of the work in this field is given in [9]. It seems however that the core issue investigated in our research which is the prediction of workflow activity failures - that may lead to compliance violations - has not been addressed before.

Domain specific modelling has gained considerable attention in the research community. Recommendations about when and how to develop domain-specific languages are given by Mernik et al. [10]. An example of such a language is the extended Compliance Rule Graph (eCRG) language [11]. An overview of domain-specific extensions of the popular BPMN modelling notation is contained in [12]. Several research teams proposed domain-specific extensions for process modelling including extensions for the modelling of clinical pathways in the healthcare domain [13] and extensions to capture specific process requirements of the maintenance management domain [14]. Conformance validation through traditional database technology has been the subject of several research teams. Snodgrass et al. proposed to store additional information in the database in order to enable a separate audit log validator [15]. Another approach is the use of Event-Condition-Action Rules. Experience with this approach for support of clinical protocols is reported in [16]. Various rule-based approaches addressing process monitoring and failure detection have been proposed. The REALM approach developed by IBM Research [17] is especially directed at compliance automation. Regulations are first expressed based on logical models and then automatically mapped into processible rules.

A comprehensive survey of online failure prediction methods and a proposal of a respective taxonomy is given in an article of a research group from the Humboldt University in Berlin [4]. In general, the failure prediction method of our work belongs to the so-called 'classifiers' that are one of several specializations of the so-called 'symptom monitoring approaches'. The classifier approaches evaluate values of system variables directly in order to classify whether the current situation is failure-prone or not. For our system approach a more refined classification scheme has been devised with the categories 'non-failure-prone', 'failure-prone', and 'highly failure-prone'.

III. Domain Specific Workflow Meta Data Model for the Detection of activity failures

Companies are often advised to address EH&S regulations by establishing a corresponding management system according to the international norm ISO 14001 [18]. A set of clearly defined processes that are oriented at a set of goals such as compliance to EH&S regulations serves as foundation of many management systems. The focus of the research reported in this article is on ICT-supported processes to enforce EH&S obligations. In particular the research is focused on three EH&S obligations that require the existence of an EH&S Regulation Management Database referred in the following by RM-DB [3]. The three obligations are: 1. the obligation to systematically establish and maintain a central registry of relevant regulations as a part of the RM-DB, 2. the obligation to carefully complete routine regulation management activities according to defined procedures (e.g. business processes and workflows, respectively), and 3. the obligation to record regulation management specific information in the RM-DB. This documentation obligation includes the recording of context information and status information about workflow activities as well as results of completed activities. A main reason for this documentation task is that through logging of activities valuable persistent data is established. This data is of high importance when internal and external EH&S audits are performed.

The above mentioned three central EH&S obligations require from companies to frequently perform EH&S regulation enforcement activities. A correct and careful completion of these activities requires to observe context-specific aspects such as specific organizational characteristics of the company (e.g. number of organizational units and decision boards). Another context-specific aspect concerns the set of relevant regulation areas (e.g. occupational safety, waste, fire, air pollution, chemical, transportation safety).

Only in an ideal world never will required activities be missed and never will they fail the required outcome. For the non-ideal real world, however, one has to consider the possibility that actually required activities will not take place and that executing activities will not lead to the required outcome. We generally refer to such situations by activity failures which may tamper an organization's efforts to enforce safety regulations with utmost reliability.

A workflow management meta data model is proposed that is specialized on the above EH&S obligations. The meta data model is directed especially at activity failures. The model considers activity failures that may occur when EH&S workflows are performed. The model is intended to serve as a foundation of an approach to detect already occurred activity failures and to predict activity failures that are likely to occur.

A concept diagram of the meta data model given in the popular Martin Notation [19] is shown in Figure 1. The boxes denote real world phenomena of the universe of discourse that possess an identity of their own. The semantic relations between the modelled phenomena are denoted by labelled edges. The concepts at the top of Figure 1 address the company specific EH&S context. The concepts at the middle layer are oriented at template data defined by modelers at workflow modelling time. The concepts at the bottom are directed at monitoring data about executing activities and also failure tracking data. One can envision that the concrete instances of the concept of Activity are created (i.e. instantiated) from the corresponding activity templates (i.e. Activity Type). The activity instances serve as proxies for real world activities that are controlled and monitored for example on the basis of a WMS.

Concepts for company-specific context data. The top part of the model in Figure 1 models the EH&S specific company context. The concepts Regulation Area, Organizational Unit,

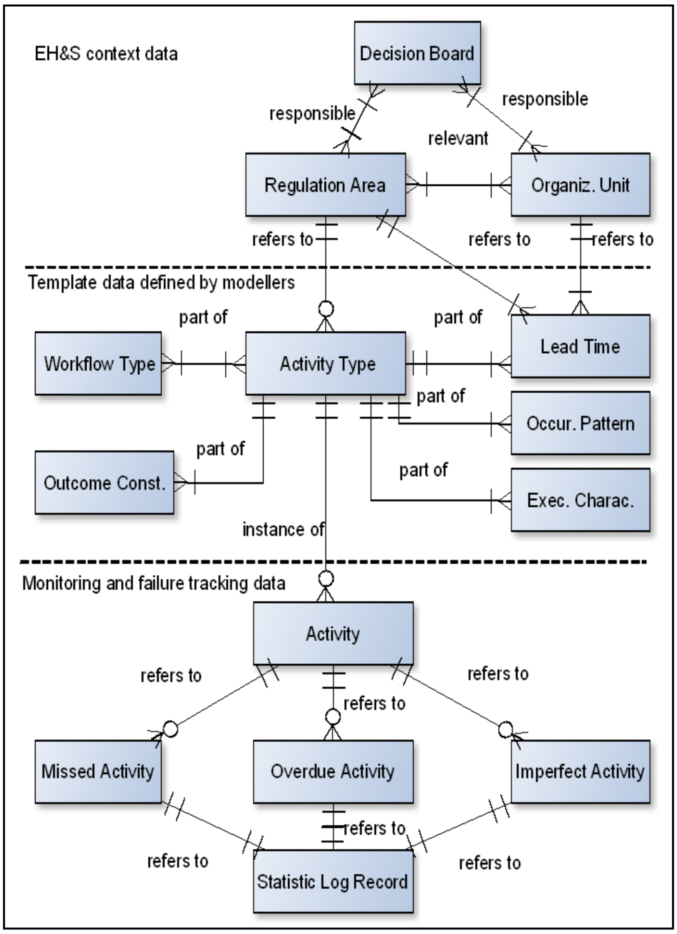

Fig. 1. Proposed workflow meta data model.

and Decision Board are addressed. That companies are usually structured into different organizational units for which different sets of the regulation areas are relevant is the first intention addressed by these concepts. The second intention is to reflect that for each relevant regulation area of an organizational unit one may assign an individual set of decision boards that are responsible for decisions in the given area. For example, the decision boards might be responsible for the selection of measures to enforce EH&S regulations [20]. One of the motivation for data modelling of the EH&S specific company context is that the data can be used to obtain indicators about the complexity of activities. It is possible to obtain from these indicators lead times of activities that are useful for failure management.

Concepts for template data defined by workflow modelers. Similar to other workflow data models the proposed model considers the concepts Workflow Type and Activity Type that serve as templates for concrete workflows and activities. That in companies with a good EH&S practice a set of pre-specified types of workflows and a set of corresponding types of activities are specified is the intention of these modelling constructs. Several modelling concepts are considered in order to model specific details of activity types as concepts of their own. The construct Occurrence Pattern reflects the occurrence characteristics of activities such as if the activity is repeatedly executed during a given time period or if the activity is triggered by a specific events. The concept Execution Characteristics reflects the execution characteristics of activities such as if the activity is completed iteratively in several steps or in an all in one approach. The concept Outcome Constraint refers to the set of conditions by which the completeness and correctness of the activity result can be validated. For most of the activities these conditions specify the set of data values to be contained in the RM-DB. The Lead Time concept is oriented at lead time specifications (i.e. minimum, average, and maximum lead time) per type of activity. Note that every lead time specification refers exactly to one particular regulation area and one particular organizational unit. Through this specification it is possible to check if given activity deadlines are met.

The specification of the lead time of an EH&S activity in the form of an educated guess requires to consider three activity-specific aspects: 1. the type of regulation area that the activity deals with, 2. the characteristics of the business activity of the referred organizational unit (e.g. is hazardous material involved in manufacturing processes), and 3. the number of involved decision boards. Consequently, per activity type a set of variants with individual lead times is considered. Every variant is associated with an individual combination of regulation areas and organizational units.

Concepts for monitoring data and failure tracking data. The concept Activity stands for activities that are performed according to the referring Activity Type. Failures that already occurred during the activity completion and failures that are likely to occur are addressed through the following three modelling concepts. The concept of Missed Activity is oriented at activity failures that emerge when an activity that is required according to its occurrence pattern has not been considered until the given deadline. That is, not even a corresponding activity instance has been created. The concept of Overdue Activity refers to individual activity instances that have been initiated according to their occurrence pattern but that missed their deadline (already). In addition, activity instances are also treated as overdue activities when these activities are likely to miss their deadline. The concept of Imperfect Activity refers to initiated activity instances that fail to meet the set of outcome constraints at the given deadline. The Statistic Log Record models a comprehensive event log about detected and predicted activity failures. The event log also contains accumulated statistical data such as the failure frequencies for the various different activity types.

IV. EXEMPLIFICATION OF THE META MODEL

Workflows models in general correspond to formal or semiformal specifications of a set of activities that serve the goal to partially or completely automate business processes [6]. To this end workflow specifications result from a refinement of business processes in terms of concrete activities and of the dependencies between activities such as temporal dependencies and input/output dependencies. The work flow specifications of our research are extended by domain-specific data, i.e. data specific to the domain of EH&S enforcement management. It is the target of this extension to establish a data foundation for the detection and prediction of missed activities, overdue activities, and imperfect activities that constitute activity failures as described above. The acquisition of the domain-specific data, for example, can be performed by a corresponding extension of the tracking and logging system component of a WMS.

On the basis of the proposed meta data model, it is possible to complement information about completed and still executing workflows and activities by further data. It is possible to use the resulting rich data in order to detect occurred activity failures and to predict activity failures. By executing proper counter-acting measures to mitigate and compensate the activity failures, it is possible to enforce EH&S regulations with a high reliability.

For the now following description of sample data for activity failure detection and prediction an essential regulation enforcement process is considered. In accordance with its main objective this considered process is sometimes referred by New Regulation Management (NRM) process [3]. The main tasks of the NRM process are: 1. to ensure that new regulations which are potentially relevant for the company are recognized, 2. to evaluate new regulations in terms of the company specific relevance, and 3. to accordingly enforce new regulations through a careful selection and implementation of proper measures. Like for all processes in the field of EH&S regulation management it is also a major task of the NRM process to comprehensively document the actions and the progress.

From industry partners we learnt that workflow modelers are advised to establish an NRM workflow that is composed of six activities [3]. These subsequent activities are:

A1: Monitor, filter, and capture new regulation. The relevant information channels (e.g. EH&S information services) of the EH&S rule setters are monitored. The announcements that pass a first rough relevance check are registered in the RM-DB.

A2: Judge the regulation relevance for the company. An evaluator judges the relevance of a new regulation for the company by assigning a relevance category to the regulation.

A3: Specify decision schedule for enforcement measures. When a new relevant regulation that requires enforcement measures is observed then a plan is defined for (collaborative) decisions about the set of required measures. Among others, one needs to specify who is in charge of the decision and when is the deadline of the decision.

A4: Organize and complete measure decision(s). A decision manager organizes and controls the completion of the decision plan.

A5: Implement set of measures. An implementation manager organizes and controls the implementation of the set of measures according to the given implementation plan.

A6: Evaluate effectiveness of measures. A reviewer checks the effectiveness of the implemented set of measures. When the review result meets given success criteria then a confirmation entry is made in the RM-DB. Otherwise, another workflow is initiated in order to deal with the problem of the failing measures.

When no new announcement of a new regulation was detected for a certain monitoring period – for example a calendar month – then only a very short version of the workflow is executed. The short version consists of the monitoring action of activity A1, the “closing” of the monitoring period, and the documentation that no new announcement was detected during the closed monitoring period. Through this approach a coherent and traceable activity documentation for all monitoring periods is established.

Recall from earlier that the proposed meta data model contains specific concepts to model company-specific context data. The modelled context data can be used to determine the complexity of workflow activities. Based on this complexity data and further data about executing workflows one can predict if activity failures are likely to occur.

The sample data used to demonstrate company-specific context data correspond to the specific characteristics of a real company referred by the fictive name CExperts [3]. For competitive reasons the real name of the company behind CExperts is not disclosed in this article. The company is a globally acting German mid-sized manufacturer of industrial alcohol, chemicals, and polymer with two different production sites in Germany. The EH&S workflows of CExperts are directed at 10 regulation areas that include water, occupational safety, waste, fire, radiation, and chemical. In the end of year 2015 the total body of regulations stored in CExperts’ corporate RM-DB comprised roughly 2000 regulations in these 10 areas from several different rule setters at all different levels (world, world region, country, state, community). Because, CExperts develops among others special chemical substances the potential enforcement measures include a) product revisions, b) infrastructure and compound revisions, c) manufacturing process revisions, d) workforce trainings and education, e) workforce instructions, f) workforce information. The EH&S organization of CExperts needs to deal with three different corporate organizational units. Every unit is assigned to each of its relevant regulation areas a set of four decision boards. The people of these four decision boards possess complementary expertise in the fields of product management, logistics and transportation, occupational safety, and quality management.

Of the above described activities of the NRM process for three activities sample template specifications are given and are explained in the following. According to our meta data model these templates result from a business process modelling and workflow specification activity. A modelling environment such as the open source environment Camunda BPM [21] which is able to derive processible workflow specifications from graphical process models can ease this activity. Since the sample templates are intended to exemplify the meta data model in the following the specifications are stated in verbal form. The data values in these verbal specifications reflect the specific characteristics of the sample company CExperts. Obviously, in a system implementation the verbal explanations are replaced by respective predefined and thus machine processible codes.

Table 1 contains the specification data for activity A1 (i.e. monitor, filter and capture new regulations) of the NRM process. The specification data of the occurrence pattern describes the conditions that trigger the execution of activity A1. As given by the sample data activity A1 is triggered when a new regulation announcement is recognized. The execution characteristics state that the activity is typically performed in a single step that requires only little time. That upon completion of activity A1 the RM-DB has to contain a description of the new regulation is specified by the outcome constraint. A general rule for the deadline of activity A1 is specified by the deadline component. The lead times specification data correspond to the minimum, the average, and the maximum lead time of activity A1.

TABLE I. DETAILS OF ACTIVITY A1- MONITOR, FILTER, AND CAPTURE NEW REGULATION.

Concept	Specification data
Occurrence pattern	Activity is started by an individual NRM workflow that is triggered by a new regulation.
Execution characteristics	Execution is typically performed in a single step that only requires a negligible duration.
Outcome constraint	RM-DB contains a description of the new regulation including the deadline for the relevance evaluation performed in activity A2.
Deadline	Activity completion is required within one day.
Lead Times	All regulation areas: 1/1/1

In Table 2 and Table 3 the template data for the activity A3 and A4 are given, respectively. The interpretation of the specification data is strait forward and thus not explicitly described in this article.

TABLE II. DETAILS OF ACTIVITY A3- SPECIFY DECISION SCHEDULE CONCERNING ENFORCEMENT MEASURES.

Concept	Specification Data
Occurrence pattern	Activity is triggered by a preceding activity A2 when measures are required to enforce a new relevant regulation.
Execution characteristics	Execution is typically performed in several steps that require non-negligible durations. The larger the number of organizational units the more complex the decision schedule to be defined and the larger the time demand.
Outcome constraint	The decision schedule that can be composed of a set of sub-decisions is fully described in the RM-DB.
Deadline	Completion is required within 3 days after the new regulation has been registered
Lead Times	All regulation areas: 1/3/5

TABLE III. DETAILS OF ACTIVITY A4 – ORGANIZE AND COMPLETE MEASURE DECISION(S).

Concept	Specification data
Occurrence pattern	Activity is triggered by a preceding activity A3.
Execution characteristics	Execution is typically performed in several steps that require a substantial duration. The more complex the decision schedule the more time is needed to complete the activity.
Outcome constraint	The decision results (i.e. measures) are fully described in the RM-DB.
Deadline	Completion deadline is given by the decision schedule.
Lead Times	Water: 4/8/15; Safety: 6/10/18; Chemical: 10/21/34

V. PROBABILISTIC FAILURE MANAGEMENT APPROACH

Today's workflow management systems (WMS) usually perform many tasks in order to execute workflows according to specifications given in the form of workflow models. This includes that for new workflow instances to be executed in the physical world, internal workflow proxy objects are created and maintained [6]. A specific WMS component referred to by "Workflow Engine" usually performs these runtime proxy management tasks. The corresponding workflow models are specified through the use of a workflow modelling environment that is often an integrated component of WMS.

Every proxy object represents and mirrors a referring real world workflow. Similarly, workflow engines instantiate and maintain activity proxy objects for the constituent real world activities of workflows. In order to enforce that the execution of workflows and activities conforms to the referring workflow models, WMS track workflows and activities in realtime and maintain a corresponding data log.

In the following it is described how the proposed workflow meta data model can be used for a new approach to predict and to detect failures of executing safety enforcement workflows. This approach draws on an extension of the traditional workflow management data log by failure management specific data items as considered in the workflow meta data model. Based on the extended logging data, a system instance is able to obtain the current failure status of ongoing workflows. The following Section A gives an overview of the major principles of the proposed approach. The major considerations for the detection and prediction of activity failures are described in Section B. Section C contains a brief scientific evaluation of the proposed approach.

A. Major Principles

In general, the capabilities of WMS to support workflow modeling and runtime execution management of proxy objects are based on a corresponding meta data workflow model. The majority of the existing workflow meta data models do not address domain specific concepts because WMS are primarily developed as general purpose systems. As opposed to that, the proposed failure management approach builds on a WMS that uses the workflow meta data model described above. That is, the concepts of the meta data model serve as basis for the specification of workflow models by the users and also for the runtime management of proxy objects (workflow instances and activity instances) by the workflow engine.

The major principles of the failure detection and prediction approach are illustrated in Figure 2. For every activity proxy object that refers to an enforcement activity, there is a comprehensive data set for failure detection and prediction supplied by the corresponding activity template. Additionally, EH&S context data specified by the workflow modeler is considered for failure detection and prediction. Also used are the statistic failure data and further logging data that are continuously maintained by the WMS. The three types of activity failures addressed in the meta data workflow model (i.e. missed activities, overdue activities, and imperfect activities) are

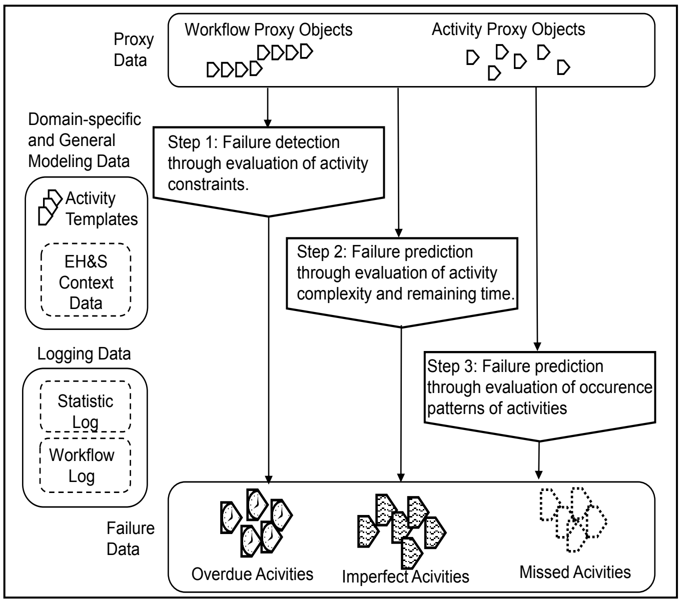
Proxy Data
Workflow Proxy Objects
Activity Proxy Objects
Domain-specific and General Modeling Data
Activity Templates
EH&S Context Data
Step 1: Failure detection through evaluation of activity constraints.
Step 2: Failure prediction through evaluation of activity complexity and remaining time.
Step 3: Failure prediction through evaluation of occurence patterns of activities
Logging Data
Statistic Log
Workflow Log
Failure Data
Overdue Acivities
Imperfect Acivities
Missed Acivities

Fig. 2. Principles of failure detection and prediction approach.

the target of the processing steps shown in Figure 2. In the next section these steps are described in more detail. The failures that are identified in the steps are indicated in the form of activity failure objects.

B. Detection and Prediction of Activity Failures

The three processing steps that are shown in Figure 2 are oriented at activity failures. Step one targets the detection of occurred (i.e. evident) failures that are imperfect activities and overdue activities. The failures are detected by evaluating the constraints that are specified in the activity templates. Obviously, the individual data available at runtime for every activity proxy are used for a corresponding constraint check. To give a concrete example, consider the above specification of the NRM process's activity A1 (*A1: Monitor, filter, and capture new regulation)*. The outcome constraint requires that a description of the new regulation has to be available when the activity is finished. When this constraint is not met, an activity failure of type imperfect activity needs to be handled. Similarly, by a comparison of corresponding proxy data with the individual activity deadline, it is possible to obtain activity failures that are overdue activities. Note that the individual activity deadlines are computed from the generally specified deadline constraint of the referring activity template.

It is the goal of step two to make use of available modelling data and runtime data in order to predict activity failures. In particular, the prediction step targets activity failures that did not yet occur but that are expected to happen in the future when no attention is paid to the potential failure cause. Figure 3 gives a high-level overview of step two using activity A1 of the workflow described in Section 4 as an example.

At first a corresponding proxy object for activity A1 is instantiated. Next, the set of failure probability indicators for the proxy object is computed based on data generally defined in the indicator formulas using the corresponding current data values.

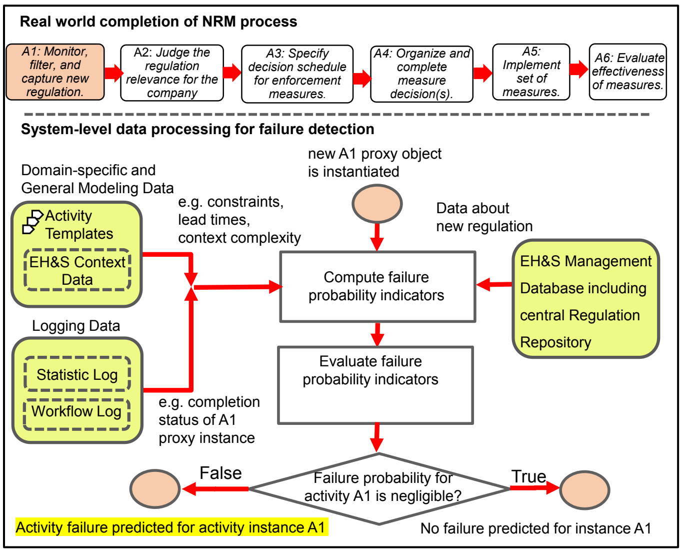
Real world completion of NRM process
A1: Monitor, filter, and capture new regulation.
A2: Judge the regulation relevance for the company
A3: Specify decision schedule for enforcement measures.
A4: Organize and complete measure decision(s).
A5: Implement set of measures.
A6: Evaluate effectiveness of measures.
System-level data processing for failure detection
Domain-specific and General Modeling Data
new A1 proxy object is instantiated
Activity Templates
EH&S Context Data
e.g. constraints, lead times, context complexity
Data about new regulation
Compute failure probability indicators
EH&S Management Database including central Regulation Repository
Logging Data
Statistic Log
Workflow Log
Evaluate failure probability indicators
e.g. completion status of A1 proxy instance
False
Failure probability for activity A1 is negligible?
True
Activity failure predicted for activity instance A1
No failure predicted for instance A1

Fig. 3. Prediction of overdue activities and imperfect activities.

This data includes modeling data of the relevant activity template such as the deadline, the complexity, and the lead times. Also logging data such as the completion status of the activity and data about new EH&S regulations are processed in order to obtain the failure probability indicators. The use of the complexity indicators reflects the general fact that usually there exists a direct positive correlation between an activity's complexity and the likelihood of activity failures. Moreover, also taken into account by the failure prediction methods are actual failure statistic data and data about the current status of the activities.

The obtained failure probability indicators serve as basis for a decision step that follows next. In general, in this next step it is focused on two questions: 1. How likely is it that the activity will be completed until the given deadline? 2. How likely is it that the results expected from the activity will be achieved? In the decision step, based on the failure probability indicators, a qualitative prediction measure is obtained. This measure determines whether an activity failure for the investigated activity instance has to be considered or not.

The activity failures that are predicted by the methods may either correspond to an overdue activity or an imperfect activity. An overdue activity is predicted when the given deadline will most likely be failed by the activity. When it is likely that the activity will not fulfill the specified outcome constraints, then an imperfect activity will be predicted.

Also in the third step a prediction of activity failures is performed targeting failures that are missed activities. The prediction method for missed activities makes use of the occurrence pattern and execution characteristics that are supplied by the activity templates. It is checked if the specified pattern and execution characteristics imply the existence of an activity. Recall that these activities correspond to real world activities that are expected to be performed. In turn, it is checked if a corresponding activity proxy object exists, indeed. When two conditions hold true, 1. no proxy object is found and 2. the

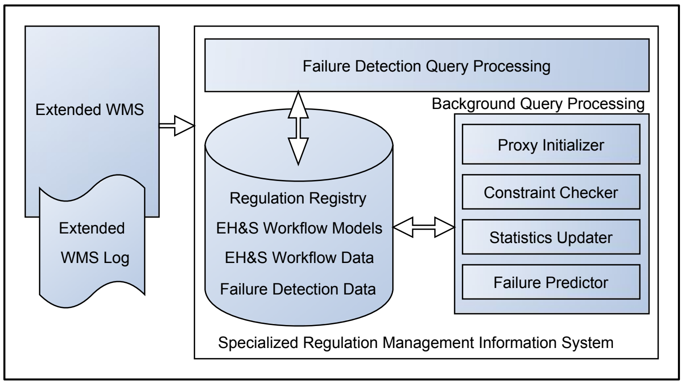

Fig. 4. High-level architecture of prototype.

temporal conditions of the activity and the workflow require the activity to be started, then a missed activity is predicted.

C. *An Initial Evaluation*

To our knowledge the approach to use a meta data workflow model that is specialized on failure detection and failure prediction in the domain of EH&S workflows is a new approach. In the next section it is described how this approach can be implemented leading to an integrated information system solution. The implementation of a corresponding research prototype is an ongoing project. It is planned to use the resulting prototype for comprehensive evaluation studies with the real world data of CExperts described in this work. The lab studies will provide validation data for our approach and insights about possibilities for improvement. Looking at the “bigger picture” of our approach, one can already at this early research phase state, that the proposed solution bears promising potential to improve the reliability of safety regulation enforcing workflows. The enforcement workflows are actively monitored with respect to their temporal constraints and outcome constraints. Also, the proposed solution provides the capability to detect missing workflows. When occurred failures are detected and failures that are likely to occur are predicted, corresponding failure data is made available for further failure handling actions. One can consider especially the failure handling action to actively provide users with alerts, failure data, and background data to effectively cope with the situation. It can be expected that this kind of active assistance being offered to safety managers for coping with failures in EH&S management tasks will promote reliability of regulation enforcement tasks.

VI. Prototype Implementation

A research prototype to demonstrate and evaluate the above probabilistic failure management approach has been devised. The prototype builds on an integrated information system that combines an extended WMS with a specialized Regulation Management Information System (RM-IS). The extension of the WMS concerns the support of workflow models that are augmented with detailed failure management data such as outcome constraints, occurrence patterns, and execution characteristics of activities. The RM-IS is specialized towards the processing and storage of failure monitoring and tracking data such as data about missed activities and overdue activities.

Figure 4 illustrates the high-level architecture of the prototype. The database stores among other data the Regulation Registry with all regulations and corresponding relevance information, the EH&S workflow models (i.e. templates for workflow instances), data about ongoing and completed workflow instances and activity proxies, and data for the detection of activity failures that already occurred or that are likely to occur in the near future. Failure detection query processing against the database is performed on request by interactive users who perform ad hoc queries. Additionally, this query processing is also triggered by scheduled query batch jobs such as the generation of failure reports.

In order to further clarify the notion of “extended WMS” consider that today’s WMS usually maintain an online log in order to track the states of ongoing workflows [19] in realtime. Based on the status information, the WMS determines and manages actions such as requests for completion of activities that are issued to workflow actors. For our approach, an extended WMS is envisioned that performs a very fine grained logging of both, workflows and workflow activities, as specified in the workflow meta data model. Especially, it is assumed that the begin time and completion time of every activity is logged through corresponding time stamps.

The architectural model defines four components that periodically update database objects for failure management purposes. The *Proxy Initializer* assigns to each new created activity proxy the corresponding set of initial values such as the individual activity deadline, the appropriate lead time values, and failure probabilities. Note that these initial values are copied from the respective activity type. The *Constraint Checker* checks the set of outcome constraints of activity proxies and reports the result in the respective activity property (oc_passed). The query set of a proxy that specifies the outcome constraints is executed until one of the following two termination conditions is reached. 1. When all outcome constraints are satisfied (i.e. all queries result to true), then the checking task is completed. 2. When the activity is completed and the query set was executed one more time after the activity completion, then the checking task is finished, too. The *Statistics Updater* component maintains data about failures stored in a failure occurrence log. The component also computes from this log statistical failure data in order to keep the corresponding attribute values of activity types up-to-date. That is, the *Statistics Updater* periodically updates the database with the latest statistical data about failures. This updating (“learning”) mechanism contributes to a proper degree of precision of predicted activity failures.

At the level of activity instances, failure prediction is performed based on a symptom monitoring approach [4]. A set of activity-type specific indicators is periodically evaluated in order to classify whether the current activity execution status is ‘non-failure-prone’, ‘failure-prone’, or ‘highly failure-prone’. The indicators include failure statistical data stored at the respective activity type (e.g. failure history) and relevant facts about the activity instance such as the complexity of the activity and the tightness of timing constraints. The resulting classification decision is reported in the respective activity instance's attributes (p_iact, p_oact, and p_mact) with values 'unlikely', 'likely', or 'highly likely'. The processing according to these principles is performed by the component *Failure Predictor*. Note when a new activity instance is created the failure probability failures are copied from the respective activity type. These values are periodically updated in every processing cycle of the *Failure Predictor* in order to reflect the evolving individual execution situation of the activity instance.

In a first implementation step the RM-IS has been implemented. The interfacing of the developed RM-IS with the extended WMS is simulated through corresponding data files. So far, the use of traditional relational database technology for the prototype implementation did not lead to any "dead ends" or extraordinary "workaround approaches". The latest versions of the popular SQL standard [20] supports language extensions by user-defined concepts such as user-defined data types and user-defined functions. For example, for several properties of EH&S activities user-defined datatypes have been developed according to the workflow meta data model. It is expected that with the further advancement of the demonstrator this SQL extensibility feature will even be more exploited.

The database tables ACTIVITY and FAILURE-LOG-ENTRY that store data about activities and failure monitoring data are described in Table IV and V, respectively. Data about activity failures are stored in the tables FAIL-OVERDUE-ACT, FAIL-IMPERFECT-ACT, and FAIL-MISSED-ACT. The outcome constraints of activities are encoded into SQL queries which evaluate if the database contains all of the data items that are required by the outcome constraints.

In order to give an overview of the proposed approach for the detection and prediction of activity failures, several sample queries are described next. For the graduation of activity failures with respect to the indication of failure occurrence four categories are used. The category "occurred" refers to evident failures that already occurred. Whereas, failure prediction results are classified into the three categories "highly likely", "likely", and "unlikely".

TABLE IV. DATABASE TABLE 'ACTIVITY'

Field	Type	Description
aid	int	Unique identifier of activity
act_type	int	Type of activity (foreign key)
deadline	date	Absolute deadline of activity
lt_min	int	Minimum lead time
lt_avg	int	Average lead time
lt_max	int	Maximum lead time
ts_start	datetime	Time stamp of start of activity
ts_end	datetime	Time stamp of end of activity
oc_passed	bool	Result of outcome constraint check performed by the Constraint Checker
ts_oc_chec	datetime	Time stamp of constraint check
p_iact	char	Probability of imperfect activity with possible values 'unlikely', 'likely', 'highly likely'
p_oact	char	Probability of imperfect activity with possible values 'unlikely', 'likely', 'highly likely'
p_mact	char	Probability of imperfect activity with possible values 'unlikely', 'likely', 'highly likely'

TABLE V. DATABASE TABLE 'FAILURE-LOG-ENTRY'

Field	Type	Description
lid	int	Unique identifier of log entry
ts_entered	datetime	Time stamp of log entry
activity	int	Activity concerned (foreign key)
fail_type	char	Type of failure with possible values overdue, imperfect, missed
fail_occat	char	Occurrence category of failure with possible values 'occurred', 'highly likely', 'likely', 'unlikely'
fc_before	int	No. of failures before the failure
fc_after	int	No. of failures after the failure
fr_before_	int	Failure rate before the failure
fr_after	int	Failure rate after the failure

The following two sample queries are directed at the detection of failures that are overdue activities and imperfect activities, respectively:

```
Select aid, "occurred"
Into FAIL-OVERDUE-ACT
From ACTIVITY
Where NOT ISNULL(ts_end) and ts_end > Deadline;

Select aid, "occurred"
Into FAIL-IMPERFECT-ACT
From ACTIVITY
Where NOT ISNULL(ts_end) and NOT oc_passed;
```

The next two sample queries predict (i.e. search for) activity failures that are highly likely:

```
Select aid, "highly likely"
Into FAIL-OVERDUE-ACT
From ACTIVITY
Where ISNULL(ts_end) and
lt_min > (deadline – Now());

Select aid, "highly likely"
Into FAIL-IMPERFECT-ACT
From ACTIVITY
Where ISNULL(ts_end) and NOT oc_passed and Now() <
deadline and p_iact = "highly likely"
```

The third query checks for overdue activities by selecting activities for which the remaining time is smaller than the minimum lead time. The fourth query predicts highly likely failures that are imperfect activities through respective conditions in the where-clause. Of the set of not yet terminated activities that did not exceed the deadline, those activities are selected, that did not yet pass the outcome constraint check. Recall that the outcome constraints of all currently executing activities are frequently evaluated by the *Constraint Checker* in parallel to the other query processing activities. When all defined constraints are fulfilled, then the Boolean value 'true' is assigned to the property 'oc_passed' of the respective activity instance. The clause 'p_iact = "high likely"' is directed at restricting the selection to activities for which a high failure likelihood was assessed by the *Failure Predictor*.

VII. CONCLUSION

The research being described in this article aims on a new domain-specific approach for the real time detection and prediction of failures of workflow activities. The context of the failures are activities to enforce EH&S regulations. It seems to be possible to achieve a more reliable enforcement of safety regulations through a timely detection and prediction of activity failures including missing activities.

The proposed failure detection methods and failure prediction methods make use of a diverse data set. This data set includes complexity indicators of activities, failure statistic data, and status data about activity proxy objects. The use of company specific organizational context data in order to obtain an indication of the complexity of activities is one of the novel ideas of the proposed approach.

A standalone prototype version of a probabilistic failure detection system is under development that follows the above described approach. The prototypical implementation builds on experience gained with the CCPro system. CCPro is a research prototype of an Environmental Compliance Management Information System [20]. Traditional relational database technology is used for the prototype in order to make sure that the proposed failure detection and prediction approach can easily be adopted by existing EH&S management systems. It appears that through user-defined functions and active capabilities such as triggers, relational database technology offers sufficient support for the implementation of the intended failure detection and failure prediction methods. In the next phase the prototype will be integrated with a WMS that supports the definition of domain-specific modelling concepts such as the YAWL system [22].

REFERENCES

[1] AberdeenGroup, "Compliance Management in Environment, Health and Safety, White Paper 6991," AberdeenGroup, Boston, MA, 2011.

[2] N. Gunningham, "Enforcing Environmental Regulation," *Journal of Environmental Law,* vol. 23, no. 2, pp. 169-201, 2011.

[3] H. Thimm, „IT-Supported Assurance of Environmental Law Compliance in Small and Medium Sized Enterprises,“ *Int. Journal of Computer and Information Technology,* pp. 297-305, 2015.

[4] J. Petts, "Small and medium sized enterprises and environmental compliance: attitudes among management and non-management," in *Small and medium sized enterprises and the environment*, R. Hillary, Ed., Sheffield, Greenleaf Publishing, 2000, pp. 49-60.

[5] B. Walker, J. Redmond, L. Sheridan, C. Wang and U. Goeft, "Small and medium enterprises and the environment: barriers, drivers, innovation and best practice. A review of the literature," Edith Cowan University, Australia, 2008.

[6] W. Van der Aalst and K. Van Hee, Workflow Management: Models, Methods, and Systems, Cambridge, MA: MIT press, 2004.

[7] M. Weidlich, H. Ziekow, J. Mendling, O. Guenther, M. Weske and N. Desai, "Event-based monitoring of process execution violations," *Proc. Business Process Management (BPM 2011), Clermont-Ferrand, France,* vol. LNCS 6896, pp. 182-198, 2011.

[8] A. Rozinat and W. Van der Aalst, "Conformance Checking of Processes Based on Monitoring Real Behavior," *Inf. Syst.,* vol. 1, pp. 64-95, 2008.

[9] M. Kharbili, A. de Medeiros, S. Stein and W. Van der Aalst, "Business process compliance checking: Current state and future challenges," *Proc. MobIS'08, Saarbrücken, Germany,* pp. 107-113, 2008.

[10] M. Mernik, J. Heering and A. Sloane, "When and How to Develop Domain-Specific Languages," *ACM Computing Surveys,* vol. 37, no. 4, pp. 316-344, 2005.

[11] D. Knuplesch, M. Reichert and A. Kumar, "Visually Monitoring Multiple Perspectives of Business Process Compliance," *Proc. 13th Int. Conference on Business Process Management (BPM 2015), Innsbruck, Austria,* vol. LNCS 9253, pp. 263-279, 2015.

[12] R. Braun and W. Esswein, "Classification of Domain-Specific BPMN Extensions," *Proc. 7th IFIP WG8.1 Working Conference, PoEM 2014, Manchester, UK,* vol. LNBIP197, pp. 42-57, 2014.

[13] R. Braun, H. Schlieter, M. Burwitz and W. Esswein, "Extending a Business Process Modeling Language for Domain-Specific Adaptation in Heathcare," *Proc. 12th Int. Conference Wirtschaftsinformatik (WI2015), Osnabrück, Germany,* pp. 468-481, 2015.

[14] M. Lopez-Campos and A. C. Marquez, "Modelling a Maintenance Management Framework Based on PAS 55 Standard," *Quality and Reliability Engineering International,* vol. 27, no. 6, pp. 805-820, October 2011.

[15] R. Snodgrass, S. Shilong and C. Collberg, "Tamper Detection in Audit Logs," *Proc. 30th VLDB Conference, Toronto, Canada,* pp. 504-515, 2004.

[16] K. Dube, B. Wu and J. Grimson, "Using ECA Rules in Database Systems to support Clinical Protocols," *Proc. Database and Expert Systems Applications (DEXA2002), Aix-en-Provence, France,* vol. LNCS 2453, pp. 226-235, 2002.

[17] C. Giblin, S. Müller und B. Pfitzmann, „From Regulatory Policies to Event Monitoring Rules: Towards Model-Driven Compliance Automation,“ IBM Rüschlikon, Switzerland, 2006.

[18] International Standards Organization (ISO), *ISO 14001:2015 environmental management system,* 2015.

[19] J. Martin, Information Engineering: Planning & Analysis, Book II, Englewood Cliffs, NJ: Prentice-Hall, 1990.

[20] H. Thimm, "ICT Support for Collaborative Environmental Compliance Management in SMEs - The CCPro Approach," *IEEE Int. Conf. Collaboration Technologies and Systems,* pp. 295-301, 2015.

[21] A. Fernandez, "Camunda BPM Platform Loan Assessment Process Lab," 2013. [Online]. Available: http://fundamentals-of-bpm.org/wp-content/uploads/2013/11/Camunda-BPM-Loan-Assessment-Process-Lab-v1.0.pdf. [Accessed 22 January 2016].

[22] W. van der Aalst, L. Alfred, M. Dumas and A. H. t. Hofstede, "Design and implementation of the YAWL system," in *Advanced Information Systems Engineering*, vol. 3084, Springer, 2004, pp. 142-159.

Heiko Thimm has been a Full Professor for Information Technology and Quantitative Methods since 2008 at the School of Engineering of Pforzheim University in Germany. In 2004 he joined Kiel University of Applied Sciences as a Full Professor for Business Information Systems after working 6 years for Sun Microsystems and SAP, respectively, as IT Solution Architect and System Analyst. From 1991-1998 he was a researcher at the German National Research Centre for Information Technology (GMD) where he was involved in various pioneering Internet and database research projects. He earned a PhD degree from the Technical University in Darmstadt and a MSc degree from New Jersey Institute of Technology both in Computer Science.

In his current research he investigates the use of recent IT technology advancements such as sensor networks, machine learning, and big data approaches for the development of next generation environmental compliance management and sustainability management information systems. He is especially focusing on supporting corporate environmental compliance and EH&S managers with active and smart assistance systems that are able to provide, among others, risk management information. Prof. Thimm is also involved in various Industry 4.0 / CPPS projects with partners from the German car/automotive manufacturing industry.