ABSTRACT
Documents are a key element in the functioning of organizations to the extent that they typically embody information. Managing and sharing documents between different stakeholders is an aspect covered by multiple Document Management and Business Process Management Solutions, many times not in an integrated manner. These two relate to each other in what concerns to the document management business process structure, where embodied the management of documents life cycle. This thesis aims to build an integrated solution of the two perspectives (Document Management and Business Process Management), applied to the context of a document management system, edoclink.

Keywords
Document Management; Business Process Management; BPM; Decision Rules; edoclink.

1. INTRODUCTION

On structured information, stored in the form of document is one of the most valuable assets in all organizations, serving as a fuel for business processes, and it is also crucial for compliance and legal retention periods [1]. Even so, there are tools able to manage the life cycle of documents efficiently and effectively. However, in what concerns to managing and sharing these documents between different stakeholders, Document Management Solutions still have a long path ahead.

Document management processes are generally mapped as ad hoc processes, that are characterized by lacking structure, and occasionally through pre-defined templates that restrict the freedom of carrying out any activity at any time or impose a set of required activities. On the other hand, there are other processes, usually called repeatable processes, whose management follows a well-structured flow according to predictable and organized communications patterns.

Still, not every documental processes have the capability to become structured, although they may present some predictable patterns. There must be maintained a certain degree of flexibility in order to allow a progressive and incremental conduction of the process according to the information provided by the business domain. Nevertheless, this certain degree of flexibility isn’t yet supported by any document management tool in an effective and efficient way.

Business Process Management is an area that involves any combination of modelling, automation, execution, control, measurement and optimization of business activities flows with the purpose of supporting the business objectives, systems involved, employees, customers and partners within and outside the boundaries of the organization [2]. Generally, BPM solutions allow to support any kind of business processes, however, managing the documents full lifecycle is not in BPM focus, as it doesn’t facilitate the understanding of how the documents evolve through time.

Therefore, finding a proper integration between document management processes and their mapping in process management tools is still a problem that this thesis aims to solve.

2. RELATED WORK

2.1 DOCUMENT MANAGEMENT

Document Management is defined by AIIM as the use of a computer system or software to store, manage and track electronic documents and electronic images captured from physical documents by a scanner [3]. In some way, a Document Management System (DMS) is responsible for controlling and organizing the document from the entire organization in a holistic way.

In the current DMS, gathering structured information from documents to be used in Information Systems is a difficult task, due to the heterogeneity of documents. Still, looking to facilitate and streamline the management of the documents, the existence of a mechanism capable of characterizing the documents accordingly to their specificities and therefore gather structured information could largely improve the efficiency of this systems.

2.2 BUSINESS PROCESS MANAGEMENT

As referred before, DMS still lack the capability of managing the collaboration associated with documents, as they do not support business processes in an efficient and effective way.

There are several classifications of business process kinds, according to multiple criteria. In this work we aim to study business processes according to the existing correlation between human intervention and structure (relationship between process execution and its formal specification).

In respect to human intervention, business processes can be classified in three groups: Person-to-Person (P2P), characterized by having primarily human involvement along its course, Person-to-Application (P2A), characterized by involving both human and computational systems and Application-to-Application (A2A) that only involve activities performed by software systems [4].

On the other hand, regarding process structure, there are four classification groups. A process is said to be unframed when there isn’t an explicit model associated with it. Ad hoc framed means that there’s a model that represents a skeleton for execution, but never the complete executed process specification. In turn, loosely framed is related to ad hoc framed processes as there is also an a priori defined model. However, in this case, there is a set of
constraints that describe the possible adaptations that can be made. Finally, and the most structured kind of process, tightly framed means that the process is executed exactly as it is defined in the a priori model [4].

The figure below depicts the mapping between the aforementioned process classifications taxonomies, where it is possible to identify an expected correlation between them. Processes that require more intensive human intervention tend to be less structured, introducing the knowledge intensive processes. On the other side, processes that require less human intervention tend to be more structured, classified as repeatable.

![Figure 1 - Processes human intervention vs structure](image)

**Figure 1 - Processes human intervention vs structure**

In Document Management context, analysing the use of edoclink in hundreds of organizations, we can conclude that we need to cover all kinds of business processes and the type to use depends strongly on their use on the regulatory framework, on the processes, and even on the knowledge of the employees. As with almost everything that interferes with the organization, the organization must be given means to select the option that best fits the business. By using reference values, 60% of the processes are better managed if they are unframed, because users know the most effective way to accomplish their task, but 40% should be preset for compliance reasons as defined by law or by the organization’s management. If we further divide the second set of processes, we would still allow partial changes to the flow (ad hoc framed and loosely framed), for example, as often happens to handle a number of cases of exception that can occur, and the remaining should be structured in a succession of well-defined steps with strict rules (tightly framed) [5].

### 2.3 CASE MANAGEMENT

In some cases, people describe processes as being tightly framed, not concerning about flexibility. Yet, several factors may lead to the need of adapting the process, requiring the existence of mechanisms to accomplish it. In addition, the number of exception cases is generally quite high and in some cases almost unpredictable [6].

Sometimes, this limitation can be resolved through the use of ad hoc treatment, described as the traditional treatment adopted by Document Management Solutions. However, in some cases, ad hoc treatment is not a very good solution as there are no restrictions regarding what each intervenient may or may not be able to perform.

Thus, the concept of case management emerges, defined by Forester as a highly structured process, yet collaborative, dynamic and guided by external events that require incremental and progressive replies according to each case. This concept can be mapped with the aforementioned taxonomy as being loosely framed processes [7].

Through this approach it is possible to manage the unpredictability associated with document collaboration in a more efficient and effective way and moreover, traceable.

### 2.4 DECISION RULES

As an outcome of the necessity to create a mechanism to automatize decision making, the decision rules arise. In fact, decision rules are a type of business rules that produce an exact determination, i.e. an exact output of what results [8], [9]. Decision rules can be described succinctly as automated decisions that can alter the course of a process.

A decision has its genesis in the analysis of a given context and is made taking into account a number of conditions. Given that the decision rules can be described as "if CONDITIONS then ACTIONS", it is then possible to map the decision-making of the documental processes in this structure by identifying the conditions that lead to a particular action. Still, this approach is only possible in decisions that can be standardized, because only in this case one can identify precisely the conditions and the corresponding triggered action.

### 2.5 edoclink

Most of the so called Document Management seek to solve only some of the problems mentioned above[5].

edoclink is an application focused on the availability of all document functions. The core is the document and its life, not just a mere document registration and search [5].

Conceptually, edoclink maps the physical document management principles. Fig. 2 depicts edoclink conceptual model.

![Figure 2 - edoclink conceptual model](image)

**Figure 2 - edoclink conceptual model**

In edoclink, a document is constituted by an assembly of one or more files associated with a set of metadata that may be generic (date, subject, comments) or specific to each document according to the business context (additional fields). The association between a document and its metadata is named record. Each record is associated to a book which enables the possibility to configure some details like additional fields and numerator to use. A numerator provides a numbering that allows to identify univocally a certain record. Additionally, given that a document isn’t static, it is necessary to maintain each document version.

Documents are a key element to organizations. Therefore, there is a collaborative process associated to each document involving multiple people. Thus, distribution is the name given to that flow, which is composed by a set of stages. To model a process, edoclink uses the concept of distribution template, that is composed by a set of template stages. So, in edoclink a business process is mapped to a distribution, if referring to an executing instance or a distribution template if referring to a modelled one.
edoclink supports manual decisions thought transitions that comprise a distribution template to be added to a distribution if this path is selected.

In order to enable more efficient access to documents related about a given subject, the concept of folder appears. It can be seen as a way to aggregate records and distributions and is particularly important because organizations need to be able to manage a lot of documents, organized by client or by theme, thus being able to access more quickly to what they need.

Just like records, each folder also provides a set of metadata, which is configured in the context of the folder classification. In order to enable the documents lifecycle, a retention plan is specified based on the classification plan according to each organization needs and using MoReq2010 standard [10].

As the subject of this work is evolving document management processes and not document management itself, security details are not described here.

3. PROPOSAL

I n a DMS, processes tend to need human intervention in most cases. Yet, there are some cases where that intervention can be minimized or even erased through automation mechanisms based on rules that control the flow of the process.

According to the aforementioned analysis and to the discussions with edoclink product team, it was possible to establish a list of requirements that the solution must fulfil in order to correctly solve the proposed problem:

R1. Cover the entire Business Process spectrum in what concerns to structure
R2. Model Business Processes according to the standard notation BPMN
R3. Automation of decision-making
R4. Structured information gathering in order to improve decision-making
R5. Ease of use and configuration for any stakeholder

3.1 COVERING ALL BUSINESS PROCESSES

As stated before, a business process can be mapped with a distribution with respect to its execution. In order to fulfil the first requirement, a mapping between the process structure taxonomy and edoclink concepts was made. Thereby, the resultant distribution structure taxonomy is:

- **Unframed distributions**
  - Distributions whose stages aren’t a result from modelling, i.e. none of the distribution stages corresponds to a template stage.
- **Ad hoc framed distributions**
  - Distributions that contain some stages that result from the instantiation of distribution templates, providing a skeleton that can be adapted accordingly to the flexibility given to each user.
- **Loosely framed distributions**
  - Distributions whose stages are defined through the progress of the distribution, based in an activities list previously defined. In reality, each activity corresponds to a sub-distribution template that is added to the distribution by the user as needed.
- **Tightly framed distributions**
  - Distributions whose all executed stages are a direct consequence of the instantiation of a distribution template, without any kind of flexibility.

Notice that both ad hoc framed and tightly framed distributions are a result of the instantiation of a distribution template. However, in tightly framed distribution, all the distribution is the result of the instantiation of a single distribution template.

Apart from the loosely framed distributions, edoclink already supports other kinds of distribution through distribution and distribution template concepts. Loosely framed distributions are not supported due to the fact that, although one can add sub-distribution templates at any point of a distribution, there is no mechanism to define, for each distribution, which sub-distribution templates can be used, for which purpose we introduced the concept of distribution type.

Along with the sub-distribution templates list, a distribution type provides the first stage configuration and an optative distribution template to be automatically added to a distribution once the corresponding distribution type is selected. Additionally, in this context it may also be possible to parametrize a set of additional fields to the distribution.

3.2 MODELLING IN BPMN

Regarding the second requirement, which reflects the need of a tool to enable the possibility of modelling processes using a BPMN notation in edoclink, two different tools were introduced: edoclink Modeller and edoclink Viewer.

These two diverge in what concerns to what is modelled/viewed: while edoclink Modeller deals with distribution templates and thereby business processes in modelling phase, edoclink Viewer is responsible for distributions execution which means executing business processes.

By introducing these tools, an alternative graphical and more intuitive configuration interface, which is in accordance with BPMN standard, will be available to all users.

3.3 ENABLING DECISION-MAKING AUTOMATION

Decision-making automation is not a straightforward requirement to fulfill. In order to automate decisions, there must be a way of specifying what are the conditions that lead to each action, i.e. decision. As aforementioned, decision rules provide a definitive and structured mechanism to specify both the conditions and actions.

However, specifying decisions is not enough to automate the decision-making process because there must be some kind of rule evaluator to enable automation. In literature, it is defended that the most effective way to enable automation is through a rule engine [9]. A rule engine is composed by three main components: user interface, inference engine and knowledge base [11].

The user interface component corresponds to the parametrizing interface that users should use to create and modify decision rules. Therefore, regarding the fifth requirement, about usability, this interface must have a high usability to facilitate the configuration by any kind of stakeholders.

The inference engine is basically where the intelligence lies. This component is responsible by evaluating the conditions in a particular context using a forward-chaining method [11].

Finally, the knowledge base component corresponds to the information that may be used by the inference engine to verify
In order to implement a business process in edoclink, there are two distinct phases: Process Analysis and Tool parametrizing.

**3.4 COLLECTING STRUCTURED INFORMATION**

During document collaborative processes, some decisions have to be made based on the current context. In previous edoclink versions, over the course of a distribution, only a free text field was collected. This lack of structuring information lead to erroneous decisions due to the fact that the information supporting decision-making was mainly unstructured and difficult to analyse.

Additionally, with the introduction of decision rules, the need to collect and maintain structured information becomes even more significant. This said, the idea of creating some kind of stage forms that embody a set of fields took shape.

It was now necessary to define how would these forms and fields be managed. Basically, form field corresponds to an additional field that may come from different contexts such as record, folder, distribution or stage. On the other hand, a form is composed by a set of ordered fields that are shown in the corresponding stage.

**3.5 ARCHITECTURE OVERVIEW**

Having discussed how each requirement would be fulfilled, the solution architecture can now be identified as the construction of a set of components which are stated in the figure below.

![Solution Architecture](image)

**4. COMPUTATIONAL MODEL**

After stating the proposed solution architecture, it is important to understand how can it be used, through the specification of a computational model.

In order to implement a business process in edoclink, there are two distinct phases: Process Analysis and Tool parametrizing.

**4.1 PROCESS ANALYSIS**

The process analysis phase comprises the following steps:

1. Identification of the stages that must be always executed
   - This step corresponds to the identification of the stages whose execution do not derive of any conditions and thus, will be part of all distributions of the process.

2. Identification of the optative paths and how can each one be activated
   - This step corresponds to the identification of the possible variations that may exist in different distributions. Note that it is important to identify the reason that leads to each alternative in order to map this optative paths as distribution templates that may be added to the distribution in three different ways:
     i. Automatically, throughout the formal definition of the conditions that lead to the activation of the respective option, parametrized in edoclink as rules.
     ii. Manual, but in a mandatory decision point between multiple alternative paths, parametrized in edoclink as transitions
     iii. Manual, according to user needs in each situation, parametrized through the list of sub-distribution templates, existing in each distribution type

3. Identification of which information is to be collected or is needed to make decisions in each stage.
   - This step corresponds to the identification of which fields will compose the form of each stage. By default, each stage only has a free text field as aforesaid in section 3.4.

4. Identification possible integrations with other systems
   - This step corresponds to the identification of integration points with other systems that may exist, which can be automatic, based in a set of defined conditions at a given moment or through their disposal to users.

5. Identification exception cases based on temporal events or happenings.
   - This step corresponds to the identification of exception cases to be handled automatically by the system during process execution.

As a result of this phase, it is possible to identify the type of process that we pretend to implement in edoclink according to its structure:

1. **Unframed distribution**
   - If it isn’t possible to identify any stage that is always executed.

2. **Ad hoc unframed distribution**
   - If it is possible to identify some of the distribution stages. Yet, it is possible to add additional stages during the execution through rules, transitions or any other sub-distribution template.

3. **Loosely framed distribution**
   - If it is possible to identify some of the distribution stages. Yet, it is possible to add additional stages during the execution through rules, transitions or a sub-distribution template from the distribution type sub-distribution templates list.

4. **Tightly framed distribution**
   - If it is possible to identify every stage of the distribution and is impossible to add new stages during the execution.

**4.2 TOOL PARAMETRIZATION**

After the first phase, the user should be able to parametrize the process in edoclink back office, starting by creating the distribution type. This step embodies the configuration of the first stage of the distribution, a possible distribution template for the following stages, a list of distribution additional fields and finally, a list of sub-distribution templates to use as needed.

The configuration of the distribution template, supported by the previous analysis is composed by the creation a subsequent configuration of each stage.
The configuration of each stage involves the configuration of the stage properties. These properties include its form, alternative paths (rules and transitions), actions and alarms.

After the parametrization of the business process in edoclink, it should be possible to instantiate it through the selection of that distribution type in the moment of creation of a new distribution.

4.2.1 Forms
A form corresponds to a configuration in the context of a distribution template stage
As aforesaid, a form enhances the possibility to collect structured information during a distribution, independently from the context of such information. Thus, each form field is associated to an additional field and to its origin context. These contexts map the various contexts where it is possible to store information based on additional fields: Folder, Record, Stage and Distribution.

However, to enable a deterministic referentiation of each field, we introduced the concept of context instances, that describe the specific object to which the field belongs (First Folder, First Record, First Stage, Current Stage and First Distribution).

For each form field, it is possible to configure, e.g., a default value, if it is editable and if it is mandatory.

4.2.2 Rules
Rules represent the decision-making automation mechanism used in this work. A rule can be seen as:

\[
\text{IF}\quad <\text{conditions}>\quad \text{THEN}\quad <\text{actions}>
\]

Where conditions can be seen as the following regular expression:

\[
(field)(operator)(value)?\quad (connector\quad (field)\quad (operator)(value)?\quad )^*
\]

A field corresponds to an additional field from a particular context. An operator corresponds to the method of comparison between the field and the configured value and a connector represents the connection between multiple conditions (AND/OR).

Each action can be characterized as:

- Distribution Template
  Insertion of a distribution template after the current stage
- Block
  Block of the distribution progress
- Action
  Service call throughout edoclink actions

4.2.3 Transitions
Transitions aren’t a new concept in this work. Each transition corresponds to a manual decision that adds a distribution template to the distribution after the current stage.

4.2.4 Actions
As aforementioned, actions allow the integration with external systems. In this way, if the necessity of creating such integration was identified, this concept should be used.

4.2.5 Alarms.
The alarms correspond to the exception cases that are treated automatically.

5. DEVELOPMENT

During the development phase, the discussed components were implemented in edoclink. Due to the fact that edoclink is organized in three layers (Figure 4), Data, Business and Presentation, it was necessary to change all of them.

![Figure 4 - edoclink layers](image)

Data layer modifications involved creating and changing some tables, views and stored procedures. In what concerns to the business layer, some new classes were introduced as well as some existing were changed in order to facilitate the fulfillment this project goals. In the presentation layer, interfaces to configure and use all the new components were also introduced as well as some services to enable programmatically interaction.

From a general point of view, the mapping between requirements and the developed components is depicted in Table 1.

<table>
<thead>
<tr>
<th>Requirements</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover the entire Business Process spectrum in what concerns to structure</td>
<td>Mechanism to support Distribution Type concept</td>
</tr>
<tr>
<td>Model Business Processes according to the standard notation BPMN</td>
<td>BPMN components edoclink Modeller and edoclink Viewer Import BPMN processes using XPDL</td>
</tr>
<tr>
<td>Enable decision-making automation</td>
<td>Data model and corresponding components in the other layers in order to support definition and use of forms</td>
</tr>
<tr>
<td>Structured information gathering in order to improve decision-making</td>
<td>Extend additional fields to use them in distributions and stages. Data model and corresponding components in the other layers in order to support definition and use of forms</td>
</tr>
</tbody>
</table>

Table 1 - Mapping between requirements and developments

An example of the modelling interface, more precisely edoclink Viewer, is depicted in Figure 5.
6. VALIDATION

The validation of the proposed approach included the implementation of various demonstrations and a pilot project.

The demonstrations have been made to potential customers through some proofs of concept which show that the solution is able to solve the existing needs. These POC’s correspond to the configuration of some business processes in the solution, using the proposed approach:

- Travelling Request
- Invoice Approbation
- Urbanistic Licensing Request
- Insurance Customer Management

The pilot project should have been done in the context of a real customer, and was scheduled for September. However, at the time of writing, it still did not start due to bureaucratic constraints related to the projects. As an alternative, a small project pilot was carried out by edoclink projects team. This project consisted in three phases:

I. Training
   Small training presentation to explain edoclink evolutions

II. Configuration
   Each element of the team configured multiple business processes in edoclink.

III. Evaluation
   Each element of the team presented their opinion in a survey about the proposed approach. This opinion was very positive which means that the objectives have been met.

7. CONCLUSION

Through the development of this work, the proposed objectives have been met. From a theoretical perspective, an effective and efficient integration between Document Management Processes and Document Management Systems was made. From a practical perspective, a solution to meet this integration was implemented and the recovered feedback was very positive.

As an output of this implementation, edoclink launched the 6.1 release that comprises this approach that is already sold to some customers.

8. FUTURE WORK

Still, as a consequence of the substantial amount of complexity that is associated with this work, additional developments that could make this solution even more complete were identified:

I. Tool for creating advanced forms
   In some cases, an ordered list of fields may not be enough to fulfil customer needs. Therefore, creating a mechanism to easily define fully customized forms would be a good evolution.

II. Resource Allocator
   During the demonstrations that have been carried out, we identified the need for the existence of a resource allocation mechanism based on defined rules.

III. Business Analytics
   Along the distribution, is now taken in a large volume of information has a value of business. By building a set of dashboards it would be possible to take greater advantage of this information.

9. REFERENCES


