Reference Modeling of Content Management Processes
Sjoerd van der Spoel
s.j.vanderspoel@student.utwente.nl

ABSTRACT
This research looks at content management processes in organizations, and how they can be supported by content management systems (CMS). To achieve this, the research in this paper is goal at creating reference models describing how content management processes should be modeled. These models in turn can be used to find which features are needed in a CMS to support the processes. To create the reference models, the first step is finding a grammar for representing the domain at hand: content management in organizations. This grammar is to be found using an ontological model that defines the basic elements in content management processes. The grammar best matching this ontology is then chosen as the way to create the reference models. Besides choosing an appropriate grammar for content management process reference modeling, this research looks into what the criteria for such a reference model are and what approach towards creating the model should be chosen. Finally, this paper provides the first steps towards a tool for finding a content management system.

Keywords
Content Management, Reference Model, Tool, Ontology, Model criteria.

1. INTRODUCTION
The internet provides an ever expanding stream of information, conveyed via all sorts of media. A large portion of the internet’s endless supply of information comes in the form of websites, owned and managed by anything from a single enthusiast to a multinational organization. The websites vary from simple static HTML-pages to complex combinations of functionality, design and information. What all websites have in common, is that they provide content in some form or way. But how is content defined?

Literally, content means “something contained by something else”. The question then shifts to what the container is. Boiko states that content is information including context. The context makes the difference between raw data, poorly understandable to humans but great for computers and information that is understandable to humans. The context comes in the form of metadata, or data about data. This metadata provides a container for the original data, making it content. [4]

An example of metadata is a label saying that the data it holds is an address or a first name. Without this label, the information contained could have been anything, with the metadata the information transforms to something understandable. The beauty of this content-container principle is that both can just as easily be stored on a computer. By tagging information with metadata, the information can even be processed by the computer, as it provides the computer with the way to treat it. A more complete definition of content is:

"Information that can be tagged with data so that a computer can organize and systematize its collection, management and publishing."

A system that provides the functionality presented in the definition above is called a content management system, or CMS. Such a system provides all the tools for handling content, with the exception of collecting the content. That is still up to the user of a CMS.

To complete the circle started at the beginning of this text, CMSs are often used to managed online content, starting with facilitating the collection of the content and ending with delivering it to the websites visitors. All these functions of a CMS are matched by processes in the organizations, and provide functionality to make executing these processes easier. Examples of processes include distributing content across multiple outlets (websites) and managing when content is published. Of course, what processes are involved in content management differs from small and large organizations on one end, and in between service, production and trading organizations on the other.

Almost every organization presents itself on the internet via a corporate or institutional website. Everyone of these websites contains information and according to the definition proposed above, that makes the information content. A corporate website through which that content is (in the most minimalistic case) at least presented is therefore a content management system.

A CMS is not limited to an online application, any system used to manage and deliver content falls under the definition. Possible non-web examples include a magazine editing tool. However, in this article the focus is exclusively on CMSs used in online media, either via the internet or via intranet applications.

2. PROBLEM STATEMENT
Obviously, a website merely presenting a few flat (X)HTML pages does not qualify for the title of management system. However, the majority of corporate and institutional websites provide more functionality than just presenting content. These sites also provide means for editing, managing and distributing content, making them far more complex software systems. The sheer amount of websites and approaches to management content means that there are hundreds, if not thousands of content management systems in use today.

A study among 58 companies investigating their content management from a business perspective revealed that the projects for setting up a CMS vary greatly in scope and focus. Budgets for these systems vary from $10,000 to millions of dollars [12]. However, the same study shows that the goals set out for these CM-projects are highly similar. Examples include creating a professional image for the organization’s website, improving both internal and external collaboration and saving money through more efficient information management.
Many of the goals found in [12] focus on either process improvement (more efficient/effective content management) including collaboration among workers or adding new services involving digital content. These goals are strongly related to the processes in the organizations, as they indicate that a CMS should improve these processes.

The question arises, if all these organizations are striving for a content management system that fits their budget and needs, can a common denominator be extracted in terms of the processes that a CMS should support the functionality of a CMS? Currently, an organization must perform an inventory of the requirements for such a system in terms of what processes it should support and sieve through the many CM systems available today. This step is where “the wheel is reinvented”, as every organization analyzes what processes they have concerning content management (and perhaps what processes they would like add or improve), and looks for functionality in a CMS that supports these functions, whereas many of these processes can be found in all organizations in the same sector (production, services etc.) and scale.

Another issue is that in the world of Content Management ever changing technologies and systems provide new opportunities. The flipside of this fast-paced world is that today’s breaking new technology is tomorrow’s old news. This makes deciding what the best CMS for an organization is more difficult, as the supply of CMS changes with new technological breakthroughs.

To summarize, the problem discussed in this research is:

Organizations share goals for their content management systems and share processes that these CMS should support, but there is no existing matching between these processes and functionality provided by CM systems. This causes organizations to keep reinventing the wheel, not only do they need to analyze their CM related processes, but also, they need to find a CMS based on the functionality needed to support these processes. Another problem involves this matching of functionality with organizational processes: As time advances, processes in the organization change (for instance creating new services/products) and, of course, the CMS change. The following section provides a research question focused on solving these problems.

3. RESEARCH QUESTIONS

A first step in providing a solution for the problem stated above is finding a way to inventoryize processes involving content management in an organization. Identifying these process leads to a structure of processes, in other words a model of the organization. An approach focused on providing a model of an organizations processes which applies to a multitude of organizations is called Reference Modeling.

A Reference Model definition focused on information systems is: A model providing recommendations for the designing of information systems, used to accelerate the development of organization-specific models.[1]

The aim of such a model therefore is providing a comparison for an organization between the processes defined in a (perhaps idealized) reference model and those in their own organization. The way in which this model is created is paramount to its success, because firstly a poorly constructed model provides poor comparison and secondly a model focusing on the wrong elements in an organization provides no useful information for the organization. Green and Rosemann suggest that a process based modeling style is the appropriate approach for modeling behavior, in combination with the fact that process modeling focuses on understanding the business processes underlying the IT [8].

All the above leads to one main research question, itself leading to several sub questions, discussed in the following sections:

What is needed to create a reference model that provides information on the quality of a CMS, without the model becoming inaccurate with time?

The main question breaks down into the following elements: creating a reference model, measuring quality of that model and finally retaining accuracy and usability of the model over time. All these elements are reflected in the following sub questions.

3.1 Creating a Reference Model

The first element of the main question is the creation of a reference model. Such a model consists at least of a number of processes involving CM and their interconnections.

The question then is, how should the processes be modeled? There are many techniques for modeling an organizations processes, but which one is right for creating a reference model focused on content management systems? As this question has no single right answer, more research is needed into the field of modeling grammars and an ontology of a CM system.

This results in the following sub questions:

What is an appropriate grammar to model content management processes in a reference model perspective?

How many reference models must be made to match different types of organizations with different types of content management related processes?

3.2 Providing Information on CM Quality

Given a reference model, the question is, how can this be put to use? In other words, how does a reference model provide information on the quality of a content management system? Here, by quality is meant the fit between CM processes in the organization and the functionality provided by the content management system.

Over the last years ERP suppliers such as SAP have come to rely heavily on reference process models in describing the functionality the software needs to provide [8]. This success of reference process modeling related to functionality provided may also apply to the CM field. All this leads the following sub questions:

What are criteria for a reference model that provides information on the comparison between business processes concerning content management and functionality provided by CM systems?

A next question is, can this information be put into a tool, so that it can be made of use to organizations:

How can this comparison be made into a tool, providing an easy way to analyze content management systems?

3.3 Avoiding Model Deterioration With Time

The usefulness of a reference model is challenged on two fronts: firstly changing business processes (like adding new services affecting CM) reduce the accuracy of the model and secondly changes in functionality provided by CMS reduce the usefulness of comparing processes with that functionality. Another question in the field of model usefulness is what type of model to create. A model can be based on the processes in existing (successful) organizations (a benchmark focused empirical approach) or on an implementation of the scientific
norms for CM business processes (a normative approach). This leads to these questions:

How can the usefulness of a reference model on content management be maintained while business processes and CMS functionality changes?

What is the best type of approach to creating reference model for the purposes of this research: a best-practice or normative approach?

4. RESEARCH SCOPE

With the problem statement and research questions now clear, it is time to shift the focus towards answering the research questions and, in doing that, attempting to solve the problem found. The questions proposed themselves provide a framework for the research that needs to be done. The research centers on three elements that need developing and researching: an ontology of content management related business processes, a grammar that can express all elements of the ontology, and finally the basis for a set of reference models that match different types of organizations. To achieve this goal, the start is setting the scope.

4.1 Content Scope

In this case, the main element influencing the scope is what content management is. Obviously, content management processes can be seen as strict as just those processes for controlling a content management system (updating and publishing) or as broad as every process in the organization involving content.

In the latter case even spreading internal documents in the organization (such as an agenda for a meeting) would qualify as a content management process: content is defined as data tagged with metadata that gives it meaning. In the example (the meeting agenda) the data is a list of items and the metadata is that the items represent things to be discussed in the meeting.

As stated above, this paper focuses on content management in online media, for example over the internet or corporate intranets. This provides a limit on the scope of what is still considered content for the purposes of this paper: The content must be made available online in some way or form, or must be able to be published online. This limitation has consequences for the content management processes that are to be considered, only those involved in handling content destined for publication fall within the scope.

4.2 Organization Scope

Besides a need to set the scope on what is considered content for this research, there is a need to limit the different types of organizations to be considered. As organizations are different in their goals and setup, so are their business processes concerning content management. For instance, content management will play a different role in a news or media organization than in a production organization. To avoid having to make endless amounts of models for different sectors of organizations, this paper focuses on production, trade, service and media organizations. This boundary is based on 3 major organizational types (production, trade, service) and is augmented with media organizations, as content is at the center of a media organizations business model.

5. CM PROCESSES

According to Boiko, content management processes fall into several categories: authoring, acquisition, conversion, aggregation and collection services [3]. Together, processes in each of these categories form the overall process before the publication of content.

The authoring and acquisition process categories involve collecting content, respectively by creating new content or by using existing sources, either external (i.e. internet) or intranet documents. Processes in the conversion category immediately follow the collection of content, and are responsible for converting the content markup language or format (if necessary). Aggregation processes involve combining and augmenting collected (and possibly, but not necessarily converted) content, by adding or removing context (metadata).

As the final category of processes before publishing, collection services processes involve supporting the collecting of content, and are comprised of CMS programs and functions.

As said before, the categories mentioned above apply to content management in the stage before publishing. In this case, publishing means making the content available to its intended audience. Examples of audiences include co-workers and customers. As a whole, the publishing services of a content management system are those services responsible for taking aggregated and converted content (the end-“product” of the pre-publishing processes) and making publications out of it. Schematically, a content management system including the different process categories is shown in figure 1.
Modeling method: “provides procedures by which a grammar can be used”. It can be argued that a language is the sum of a grammar and a modeling method.

The application of a grammar using a specific modeling method results, according to Wand and Weber’s definition, in a *script*.

**Script:** “Each script is a statement in the language generated by the grammar”. Fettke and Loos add to this definition: “A script is a representation of a real-world domain using a particular grammar”. Domain is used in this context as the part of the world the modeler wants to represent.

The Wand/Weber definition of the result of representing the domain using a grammar is what most often is called a model. In this research, the definition of script will be used in parallel with the more usual term *model*. A final definition, put forward by Fettke and Loos[7] is:

**Reference model:** “A script representing a class of real-world domains”. Here, the power of the reference model is brought forth, explaining that the reference model is applicable to a multitude of domains, all domains sharing a category and the properties belonging to that category.

### 6.2 Ontology elements

Before a reference model can be made, an ontology must first be composed. This ontology then serves as a basis for selecting a grammar, since it provides a way of checking if a grammar can depict what is to be modeled. The first step now is finding the elements in the ontology.

In the aforementioned map analogy the ontology was compared with the legend of a map. One difference between legend and ontology is that a legend links graphical elements on the map to a textual description of that element. An ontology only does the latter, the chosen modeling language provides for the graphical aspect.

According to Green and Rosemann[8], a modeling grammar can be judged using an ontology, leading to an insight on the strengths and weaknesses of a conceptual grammar. This judgment is based on two criteria. The first criterion, *ontological incompleteness*, occurs when elements from the ontology are not represented by at least one element of the grammar. The second criterion, *ontological clarity* specifies the extent to which the grammar maps to the ontology, i.e. how many language elements map to a single ontological element. This, of course, relates to the definition of a grammar, as its ontological quality is a consequence of how well constructs in the grammar are reflected by the ontology and vice versa.

In their work, Green and Rosemann use the Bunge-Wand-Weber (BWW) models as an ontology, and use it among others to judge several process modeling languages. This BWW models [22] provide a set of core phenomena in the information systems world, and is therefore not exclusively focused on processes. However, Green and Rosemann have shown the models to be of use for evaluating process grammars. Since the reference model this research aims to develop is a business process model, the BWW models will also be applicable.

### 6.3 Bunge-Wand-Weber models

The BWW models are based on a very generic description of the elements of an information system, recognizing things, states, events and properties for all these elements [6]. In the BWW models these basic constructs have been added upon to make a basic ontology for describing information systems in general. Obviously, as an ontology for a (CM) business process the more important elements are states and events, as a process is series of activities. Things of course also come into play, as the processes affect things. All this is generic, and can be applied to everything. This is of course also a strength underlying the BWW models, making them applicable as a tool for judging ontological quality of any grammar. However, for this research, it is interesting to know how the BWW models can be specialized so it tells more on process modeling languages in general, and the content.

As the attentive reader will have noticed, Wand and Weber developed multiple models [22] [24] [23] based on the work of Mario Bunge in the field of ontology [6]. From his work, the basic ontological elements as mentioned above (thing etc.) emerged, later put to use in the models by Wand and Weber. These models, amongst are made up of a *representation model*, a *state-tracking model* and the *good decomposition model*. The first of these focuses on the elements that make up the domain, the second concerns the states elements can experience. The final model describes systems and how they can be decomposed into subsystems. Also, the good decomposition model provides tools for assessing how good a decomposition is. This however is beyond the scope of this research.

In [8] the representation model is used as starting point for an ontological evaluation of process modeling grammars. As this research is centered on business processes, process modeling grammars are the tool of choice for representing the domain concerned, content management. Since this research is concerned with representing processes in content management, the sub-system based decomposition model and the state centered model are considered not relevant. Furthermore, the representation model has often been used to evaluate grammars [8,7,27].

With the BWW model best suited as a base ontology for this research decided, now we can look at where this model can be specialized to better fit the world of content management processes in organizations.

### 6.4 BWW Representation Model constructs

Green and Rosemann[8] extracted the following list of constructs from Bunge[6] philosophy as their ontology for checking the quality of a grammar: THING* (associate/composite), PROPERTY*(in general/in particular/ hereditary/ emergent/ intrinsic/ non-binding mutual/binding mutual/attributes), CLASS, KIND, STATE*, CONCEIVABLE STATE SPACE, STATE LAW (stability condition/corrective action), LAWFUL STATE, SPACE, CONCEIVABLE EVENT SPACE, TRANSFORMATION*, LAWFUL TRANSFORMATION (stability condition/corrective action), LAWFUL EVENT SPACE, HISTORY, ACTS ON, COUPLING (binding mutual property), SYSTEM, SYSTEM COMPOSITION, SYSTEM STRUCTURE, SUBSYSTEM, SYSTEM DECOMPOSITION, LEVEL STRUCTURE, EXTERNAL EVENT, STABLE STATE*, UNSTABLE STATE, INTERNAL EVENT, WELL-DEFINED EVENT, POORLY-DEFINED EVENT.

In this list of ontological constructs, the fundamental/ core constructs as developed in the philosophical work of Bunge[6] are marked with an asterisk (*). The elements in between parentheses after a construct represent variants on the constructs, specializations of the construct or properties of the construct. For the sake of brevity, not all of these constructs will be discussed here, but certainly the fundamental constructs deserve some clarification, as these underlie the entire ontology.
A final part of the ontology that needs explaining is a matter of definition: within this ontology a class refers to several things sharing a single property, whereas the definition of class used elsewhere in this paper refers to what is known as a kind in the BWW ontology. A kind is defined as several things sharing multiple properties.

### 6.5 Limitations of the BWW-ontology

The BWW-ontology has become a widely accepted as a modeling theory. However, Rosemann and Wyssusek suggest [16] that the model is not expressive enough. Amongst other things, they suggest the BWW-ontology is less suited for modeling more abstract domains, in other words, it’s at its best representing concrete things like materials. For, for instance, social relations the BWW-ontology is not expressive enough, as Rosemann and Wyssusek argue. The BWW-ontology not being directly applicable to factual objects, in others words, the real world.

The lack of application to factual objects is itself a result of how Wand and Weber adapted the original philosophical study by Bunge. It is beyond the scope of this research to discuss this in depth, here it suffices to say that the constructs in the Bunge-Wand-Weber ontology do not reference real objects, they are fictitious, making them by Bunges definition less applicable to empirical research (which is based on facts).

Although the above does suggest valid improvements for the Bunge-Wand-Weber ontology, these are not of importance for the purposes of this research. The ontology in this research is used as a tool for assessing the suitability of a grammar for modeling content management processes. What Rosemann and Wyssusek suggest as improvements is more applicable to assessing an organizations structure.

### 7. SELECTING A GRAMMAR

As an abundance of research has been done on ontologically evaluation grammars with the BWW-model [8,15,21], checking among others the ontological quality of the Business Process Modeling Notation (BPMN)[26] and Event-Driven Process Chains (EPC)[18], part of the ARIS architecture and of the Unified Modeling Language (UML) [5]. The latter was the most dominant process modeling grammar at the end of the last century [20], and is together with BPMN considered as the mainstream of business process modeling [27]. These grammars are therefore considered prime candidates for the modeling of the subject of this research. UML is excluded, as this grammar has no specific focus on business processes.

#### 7.1 Comparing EPC and BPMN

Green and Rosemann[8] performed an ontological evaluation of Prof. Scheer’s EPCs, the central part of the so called Architecture of Information Systems (ARIS)-house. Their research focuses on answering two questions: Is the grammar ontologically complete (see above) and is the grammar ontologically clear? These two properties for consideration are visualized in figure 3.

![Figure 2: Mapping grammatical to ontological constructs](image)

The figure shows the four situations that can occur when comparing a grammar to an ontology, the top row describing grammatical shortcomings, the bottom row showing situations where the grammar is richer (can describe more) than the ontology. Obviously, either situation is ideally avoided in a grammar.

In their study on the evolution of grammars Rosemann et al. [15] performed a comparison of event-driven process chains and the business process modeling notation (amongst other), based on the set of ontological constructs by Wand and Weber. An extract of their comparison is shown below (table 2). On a side note, their comparison leads to the conclusions that other grammars than EPC and BPMN are even more complete, but this paper considers the broad support for these two languages to be more important. The reference model should be understandable to as many process modelers as possible to maximize its potential.

### Table 1: Explanation of constructs in the BWW ontology

<table>
<thead>
<tr>
<th>Ontological construct</th>
<th>Explanation[8]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thing</td>
<td>The most elementary unit, what the world is made up of.</td>
</tr>
<tr>
<td>Property</td>
<td>A property is possessed by a thing, mapping the thing through a function into a value. Properties can be general (for all things/all things of a category), - particular (for a specific thing) - hereditary (a property of a component thing derived from one of its composite things), - emergent (a property of a component thing not derived from one of its composites), - intrinsic (belonging to individual things), - mutual (possessed by paired multiple things, either binding (saying something about the individuals in the pairs) or non-binding, only saying something about the pairing/the things in relation to each other) - attributes are the names of properties of things.</td>
</tr>
<tr>
<td>State</td>
<td>The vector of all values for all property-functions of a thing</td>
</tr>
<tr>
<td>Transformation</td>
<td>A mapping from one state to the other, describing how the things properties change.</td>
</tr>
</tbody>
</table>
Table 2: Comparison of grammatical and ontological constructs

<table>
<thead>
<tr>
<th>Ontological construct</th>
<th>Present in Modeling grammar:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPC</td>
</tr>
<tr>
<td>THING</td>
<td>√</td>
</tr>
<tr>
<td>CLASS</td>
<td></td>
</tr>
<tr>
<td>KIND</td>
<td></td>
</tr>
<tr>
<td>PROPERTY</td>
<td>√</td>
</tr>
<tr>
<td>STATE</td>
<td></td>
</tr>
<tr>
<td>CONCEIVABLE STATE SPACE</td>
<td></td>
</tr>
<tr>
<td>STATE LAW</td>
<td></td>
</tr>
<tr>
<td>LAWFUL STATE SPACE</td>
<td></td>
</tr>
<tr>
<td>STABLE STATE</td>
<td>√</td>
</tr>
<tr>
<td>UNSTABLE STATE</td>
<td></td>
</tr>
<tr>
<td>HISTORY</td>
<td></td>
</tr>
<tr>
<td>EVENT</td>
<td>√</td>
</tr>
<tr>
<td>CONCEIVABLE EVENT SPACE</td>
<td></td>
</tr>
<tr>
<td>LAWFUL EVENT</td>
<td></td>
</tr>
<tr>
<td>SPACE</td>
<td></td>
</tr>
<tr>
<td>EXTERNAL EVENT</td>
<td>√</td>
</tr>
<tr>
<td>INTERNAL EVENT</td>
<td>√</td>
</tr>
<tr>
<td>WELL-DEFINED EVENT</td>
<td>√</td>
</tr>
<tr>
<td>POORLY DEFINED EVENT</td>
<td></td>
</tr>
<tr>
<td>TRANSFORMATION</td>
<td>√</td>
</tr>
<tr>
<td>LAWFUL TRANSFORMATION</td>
<td>√</td>
</tr>
<tr>
<td>ACTS ON</td>
<td></td>
</tr>
<tr>
<td>COUPLING</td>
<td></td>
</tr>
<tr>
<td>SYSTEM</td>
<td></td>
</tr>
<tr>
<td>SYSTEM COMPOSITION</td>
<td></td>
</tr>
<tr>
<td>SYSTEM ENVIRONMENT</td>
<td></td>
</tr>
<tr>
<td>SYSTEM STRUCTURE</td>
<td></td>
</tr>
<tr>
<td>SUBSYSTEM</td>
<td></td>
</tr>
<tr>
<td>SYSTEM DECOMPOSITION</td>
<td></td>
</tr>
<tr>
<td>LEVEL STRUCTURE</td>
<td>√</td>
</tr>
</tbody>
</table>

Representational coverage: 11/29 19/29

The above table clearly shows the BPMN to be more ontologically complete, as more constructs from the ontology are mapped to the grammar. Also, both BPMN and EPC are ontologically incomplete. Based on the table, no conclusion can be drawn on whether or not either language has redundancy, grammatical excess or overload.

Recker et al. [14] suggest that BPMN is redundant, as constructs in the ontology are mapped to multiple grammatical constructs, leading to ambiguity. Also, Rosemann et al. [15] conclude that the lack of constructs mapping to the state elements in the ontology may lead to difficulty representing business rules. This last problem is not present in the EPC grammar.

Green and Rosemann[8] do state that EPCs is ontologically unclear, as some grammatical constructs map to multiple ontological constructs. Also, ontological excess exists, as connectors like and, or and exclusive or are not present in the BWW ontology.

To conclude on which language is to be used, it’s consider that ontological completeness is more important than ontological clarity. After all, the first gives an indication of how well a grammar is capable of modeling a domain, the latter just describes how understandable/unambiguous the model is. Ambiguity and understandability are of course also a function of experience using a model. This research assumes that, since BPMN is a popular process modeling grammar, the understandability argument does not play a role here. Therefore, BPMN will be the grammar used for developing the reference models in this research.

8. CRITERIA FOR REFERENCE MODELS

With an appropriate grammar selected for the purposes of this research, the core of this research comes into view: developing a reference model. Before any reference model can be made, first some boundaries need to be set, criteria to which a reference model should conform. By explicating these criteria, it is clear how a reference model can be made and how some judgment can be made on its quality. Amongst others, the question needs answering if the reference models to be constructed should be normative, empirical or based on best-practices. The criteria are divided into two (possibly overlapping) categories: general business process modeling criteria and reference modeling specific criteria.

8.1 Business process modeling criteria

Becker et al. [2] suggest guidelines for business process modeling, which are applicable to this context: modeling content management related business processes. The guidelines they propose are centered on correctness, relevance, economic efficiency, clarity, comparability and systematic design. An explanation of the guidelines can be found below:

**Correctness:** The extent to which the model complies to the syntactic and semantic rules stated by the modeling grammar and the correct use of terminology.

**Relevance:** Use of a modeling grammar fitting the domain modeled and creating a minimal model: eliminate model elements that do not add meaning. Furthermore, relevance is concerned with making sure that modeling a domain is useful, that it fulfills a need. The model created in this research is assumed to be relevant, since a) the grammar is appropriate (see above) b) a generic model like a reference model will always be minimal, with more meaning being added by specializations c) the reference model fulfills a need: it addresses the problem stated at the start of this paper.

**Economic efficiency:** Whether or not the benefits of creating a model outweigh the costs of modeling, which can be considerable. The research proposes a
The use of a common resource is important to consider in all concerned with resource sharing across multiple activities. This is strongly linked to the choice of modeling grammar, more specifically to grammatical construct overload. As stated above, BPMN has little or no construct overload. Furthermore, as BPMN is a mainstream modeling technique, a large group will find a model created using that technique understandable.

### Clarity

The understandability of the model. This is based on theoretical norms on content management processes, not the organizational information systems. The guideline is therefore considered optional.

### Comparability

Demands that choices made in the modeling project are consistently used. In other words, this guideline is concerned with following the rules set for a project for all models made in that project.

### Systematic design

This guideline postulates that a model in a process model should also be found in a data model. More generally, this guideline demands that the modeler looks at the system as a whole. For the purposes of this research, this is probably not applicable, as the focus is business processes, not the organizational information systems. The guideline is therefore considered optional.

Altogether, following the above requirements should lead to a qualitative and usable model. The above guidelines serve as a way to frame the model to be developed. All guidelines represent things to consider while (and on completion of) modeling but do not say anything on the contents of the model.

#### 8.2 Reference model specific modeling criteria

Criteria from the list above that is especially important for reference modeling is economic efficiency and relevance. Since a reference model is likely to be more complex than a domain-specific business process model, it will be more costly to develop. If more costs are made, it's obviously important to insure that the extra effort of creating a reference model pays off. Another important point is relevance, making sure that the model can actually be used for a class of domains.

As Malone et al. [10] state in their research on creating process patterns (an approach similar to creating reference models), two elements are important for a reference models: specializability and managing dependencies. The first refers to the model being suitable to expend to fit one or more specific domains more closely (reduce abstraction). Managing dependencies is specific for creating process patterns, as it proposes three different dependency-relations between processes: flow, fit and sharing, all concerned with resource sharing across multiple activities. The use of a common resource is important to consider in a reference model, but is covered by the BPMN modeling grammar. The grammar provides constructs for common research use, forcing attention to resource dependencies.

#### 8.3 Best-practice vs. normative approach

Also a property of the model is the approach chosen. Either the model is based on what successful implementations of content management processes are, the benchmarking approach or the model is based on theoretical norms on content management, the normative approach. Obviously, these approaches are quite different. Below (table 3) is a comparison of both approaches.

<table>
<thead>
<tr>
<th></th>
<th>Best-practice (empirical)</th>
<th>Normative (theoretical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on continuous research of processes in a multitude of organizations</td>
<td>Based on theoretical norms from literature, formally justified hypotheses of what the best practice will be.</td>
<td></td>
</tr>
<tr>
<td>Time-consuming</td>
<td>Arduous</td>
<td></td>
</tr>
<tr>
<td>Ground in actual business</td>
<td>Ground in literature</td>
<td></td>
</tr>
</tbody>
</table>

Some authors see a reference model as nothing more than an abstraction of collected organizational best-practices [17,13]. This is of course a valid way of combining knowledge on content management to create a reference model. The major bottleneck is then analyzing processes within many organizations, to be able to collect enough information to extract abstract and general rules, needed for a reference model.

Since the time available for this research is very limited, the best-practice approach is to be dismissed for lack of attainable sources (analyzing organizational best practices is considered too time consuming for this research). The approach to be used in this research is therefore a normative one, although best-practices found can of course still be used as a means of validating the models created. Should further research be made into the field treated in this paper, it is worth investing in gathering best-practices.

#### 9. Preventing model deterioration

An issue concerning modeling in general is that the domain models changes with time. Of course, this issue also concerns the domain covered by this paper: content management. The issue leads to the model being a poorer reflection of reality: model deterioration.

As one of the criteria for modeling a business process, economic efficiency of the model, is covered by creating a model that can be used more than once, it is obvious that tackling model deterioration is an important issue. Having a model that becomes inaccurate after a short time strongly reduces the benefits from the model.

The most simple solution for this problem is abstracting away from the domain. The more abstract the model, the higher the amount of domains covered will be. However, abstraction also increases the amount of specialization that needs to be done to make the reference models of use in specific organizations. Therefore, abstraction and postponed deterioration must be carefully balanced.

Another way of preventing model deterioration is actively keeping it up-to-date, by checking if it still conforms to (both) scientific norms and best practices. Obviously, this takes time and money.

A final way of avoiding deterioration is part of the approach chosen in this paper. By focusing on business processes, not content management system functions, these ever changing software environments are taken out of the equation, in other words, the chosen domain (business-side of CM rather than software) is less changing.

#### 10. Creating reference models

Everything needed to create a reference model has now been collected: an appropriate language, a set of criteria, an approach towards creating the model and insights on maintaining a good representation of the domain over time. The only element missing now to create reference models is a set of norms
relating to content management systems. Some of these norms can be found in [4,3].

As mentioned before, content management processes can be ordered into five separate groups or categories. The scope of this paper states that there are four types of organizations to be modeled. This leads to the conclusion that, considering that every category of processes is to be researched for every type of organization, a maximum of twenty process groups are to be modeled. However, this number may decrease if organizations across different types (for instance trade and service organizations) share norms for (certain) content management processes, thus creating a higher abstraction level in the model than can be anticipated beforehand.

10.1 Limitations on the models
As the models are developed, it becomes clear that this paper can only provide very basic models for content management. This is due to (as discussed above) a lack of information on actual processes in the different sectors (media, production, trade and services) from a normative perspective. This obviously reduces the specialization of the model, thus reducing its usability. Furthermore, the criterion of economic efficiency is jeopardized by a higher abstraction level of the models.

10.2 Authoring process
The authoring process is one of the more abstract processes in the entire process from content creation to publication. The basis of this process is “an author creates content from scratch and submits it into the system”. It is assumed this process is similar for each of the four organization types, as it is a simple process [19]. The BPMN representation of the authoring process is shown in appendix A.

10.3 Content acquisition process
The second category of processing leading to new content is the acquisition process. This process is aimed at retrieving information from external sources. As content is information combined with metadata [4], this process is to a large extent concerned with gathering that metadata (such as the name of the source, the title, the date and time of creation and description)[11]. This process has different implementations for media organizations as compared to trade, production and service organizations, as this is a core business for media. A model for content acquisition in media organizations is shown in appendix B.

The major difference between media organizations and other organizations is assumed to be the extent to which acquisitioned content must be checked. A media organizations quality of work is after all the quality of the content they offer to their audience. This research assumes improperly validated (“sources checked”) content cannot be submitted. A model for other organizations is also in appendix B.

The model for non-media organizations is a hybrid between the media acquisition model and the authoring model, under the assumption that content that is not critical for the business may still be acceptable, even if the sources, source information or metadata may not be up to the standards of the organizations.

10.4 Conversion process
The conversion process is responsible for taking the information from the content objects gathered and transforming them to a different markup language, if necessary. This transformed content serves as a basis for input into the web based content management system. Again, based on scientific norms [3,19,11] nothing can be said on the difference in this process for different organizations. A conversion process model is shown in appendix C.

10.5 Aggregation process
After the optional conversion of the content created or acquisitioned, the multiple content objects can be combined. The goal of this process is combining several sources into a uniform single content object. This involves, amongst other things, matching the metadata of the content with some metadata system provided by the content management system. A generic aggregation process can be found in appendix D.

10.6 Publication process
The final process in a cycle of creating content (the content lifecycle) is making the content available to its intended audience. As the scope of this research only includes web content management, the mechanism of making the content available is through a web interface. The publishing process is therefore concerned with spreading the content (if needed) across different web servers and web pages. Also, the process is responsible for making the content available exactly on time. This latter responsibility is important for instance in news publishing. Appendix E shows the publication process model.

10.7 Overview of content management processes
An overview of all the processes mentioned above, and how they are interlinked is shown in appendix F.

The model shows the possibility of parallel multiple content acquisition and content creation processes, which are ultimately combined via the aggregation process.

11. REFERENCE MODELS AS A TOOL FOR CMS SELECTION
A final aspect of the reference model is making it of use. Ultimately, the processes described in this research are aimed at improving content management processes within organizations, by conforming them to normative standards or standards provided by best-practices. Since these processes do not stand alone, there is also a content management system needed.

From the processes described in the appendices features for content management systems can be extracted. For example, from the authoring process it can be extracted that the supporting system should allow for content revision. This approach can be used for every aspect of the processes in the model, leading to a set of features needed to support the best possible content management process. Kang et al. argue that such a feature-centered approach towards software selection is promising [9].

Of course, not only the process-based features are important, there are other functions a CMS should support: planning, commissioning, annotation, query, retrieval [11]. With all desired functionality for the system supporting their content management processes charted, an organization can visit for instance the CMS Matrix to find a matching Content Management System. Of course, this functionality mapping could be added to the reference model, thus creating a tool.

12. CONCLUSION AND FURTHER RESEARCH
12.1 Problem & Goal feedback
This research has been an attempt of reducing how often “the wheel is reinvented” when it comes to finding a content management system. The approach chosen to tackle this problem was making a general model of how organizations should shape their content management processes. A next step then is finding a content management system that covers all the functionality needed to support the CM processes. In that sense,
this research provides a first step towards solving the main problem, not a complete solution. To achieve this, more research is needed into the norms or best practices for CM processes. However, all research questions have been answered, thus satisfying at least that goal of this research.

12.2 Answers to research questions
Below, the research questions are revisited, with an answer given to every question, where possible.

What is needed to create a reference model that provides information on the quality of a CMS, without the model becoming inaccurate with time?

To get to a general reference model for organizations, three ingredients are needed: a grammar for modeling, a set of criteria for assessing the model and an approach upon which to base the model.

How can the usefulness of a reference model on content management be maintained?

Inaccuracy over time is reduced by keeping the model abstract and by updating the model from time to time.

What is an appropriate grammar to model content management processes in a reference model perspective?

The grammar chosen to model is the Business Process Modeling Notation, as this best conforms to the ontology devised by Bunge, Wand and Weber. By matching the ontology, it is proven that the grammar can provide a proper representation of the domain modeled. In this case, that domain is one of business process modeling.

How can this comparison be made into a tool, providing an easy way to analyze content management systems?

By analyzing the developed reference model, a set of features needed for a CMS can be extracted. This feature set can be compared with existing CMSs.

How many reference models must be made to match different types of organizations with different types of content management related processes?

For the purpose of this research, four domains have been selected to model: service, trade, production and media organizations. As content management processes fall into five separate categories, a maximum of twenty models are to be made. Because not every category of CM process is different for all domains, the actual amount of models is lower.

What is the best type of approach to creating reference model for the purposes of this research?

For the approach chosen there were two options: a normative approach or a bench mark based approach. The latter is more grounded in reference modeling research, but was abandoned due to time constraints. This has lead to a less usable model, as scientific norms where found to be unspecific. The model is therefore more abstract, leading to more need for (later) specialization.

12.3 Further research

In this research, the main bottleneck was the lack of information from real organizations as to how they shape their content management processes. A proposition for further research is therefore adding best-practice information to the models, making them less abstract and so more easily applicable for a specific organization.

Besides models better grounded in the real world of content management in organizations, another item of interest is creating a tool for combining model and content management system. If an organization can see at a glance which CMS fits the idealized processes matching their type of organization, this greatly reduces the effort they need to put into finding a CMS.

REFERENCES


Appendix A. Content authoring process model
Appendix B. Content acquisition process models

The acquisition process for media organizations.
The acquisition process for non-media organizations.
Appendix C. Content conversion process models

The content conversion process for organizations in general.

Appendix D. Content aggregation process

The content aggregation process for organizations in general.
Appendix E. Content publication process

The content publication process for organizations in general.

Appendix F. Overview of content management processes

An overview of processes involved in the content management lifecycle.