Towards the Specification of Digital Content - The Enterprise Content Modeling Language (ECML)

Jan vom Brocke  
*University of Liechtenstein, jan.vom.brocke@uni.li*

Jörg Becker  
*European Research Center for Information, becker@ercis.de*

Alexander Simons  
*University of Liechtenstein, alexander.simons@hochschule.li*

Stefan Fleischer  
*University of Muenster, stefan.fleischer@ercis.uni-muenster.de*

Follow this and additional works at: [http://aisel.aisnet.org/amcis2008](http://aisel.aisnet.org/amcis2008)
ABSTRACT

Enterprise Content Management (ECM) is requisite technology for issues of compliance, governance, and business process efficiency. However, the implementation of ECM systems (ECMS) confronts organizations with a manifold of challenges. Particularly the multiplicity and diversity of digital content captured within an organization makes a tailored ECMS adoption difficult. In this paper we state that conceptual models may serve as a means for the structured analysis and design of ECMS. On the basis of a literature review, we derive specific requirements for content modeling and show that these requirements are not met by conventional modeling languages. We therefore introduce ECML, the Enterprise Content Modeling Language, which is specified by means of meta-modeling and illustrated on the basis of an exemplary application.

Keywords

Conceptual Modeling, Content, Content Model, Content Specification, Enterprise Content Management (ECM), Enterprise Content Modeling Language (ECML), Modeling Language.

INTRODUCTION

Enterprise Content Management (ECM) is an emerging field in Information Systems (IS) research dealing with strategies, processes, skills, and technologies providing the means for an efficient management of digital information (cf. Munkvold, Päivärinta, Hodne and Stangeland, 2006, p. 70; Smith and McKeen, 2003, p. 647). The Forrester Research Group expects that the demand for ECM software will reach $3.9 billion and above by the end of this year. This figure exceeds that of the overall software market, whose compound annual growth rate stands at 19%. Particularly for large-scale organizations, ECM is “the requisite technology for addressing their compliance, governance, and business process efficiency needs” (www.forrester.com). Despite its undisputed relevance, the interest in ECM mainly stems from practitioners—in contrast however, scientific interest is still disproportionately low (Tyrväinen, Päivärinta, Salminen and Iivari, 2006, p. 627). Thus, methodical support for adopting, implementing, or applying ECM systems (ECMS) can hardly be found within IS research literature (cf. vom Brocke, Simons and Cleven, 2008). In this paper we state that there is in particular a lack of suitable guidelines for specifying an organization’s individual content situation. The reasoning for this perception can be seen as follows.

As organizations create huge amounts of digital content, an adequate specification of content assets existing within an organization is crucial for successfully adopting ECMS. The IDC GROUP holds that 1,288 x 1018 bits—or 161 billion gigabytes—of digital information were captured in the year 2006 and estimates that this number will increase more than six fold by 2010 (Gantz, 2007, p. 1). Furthermore, there are lots of different content types also requiring different ECMS support (vom Brocke and Simons, 2008). For example, a construction drawing (usually created and edited by multiple users) may require ECMS functionalities supporting the process of collaboration (e.g., access rights, versioning). In contrast, a standardized invoice is created more or less automatically and may have to be archived for a certain period of time due to legal restrictions—by further assuring that it is not modified during that time. Concluding, the diligent analysis and handling of this information flood is highly important for ECM success, but however a very complex task. An ECMS implementation is further complicated as the ECMS adopters do not know about the specific managerial requirements of certain content types, whereas, vice versa, the content users are usually not aware of available ECMS functionalities. Particularly conceptual
models may serve as a means for comprehensively assessing, describing and illustrating an organization’s content situation in order to enable a sound communication between ECMS users and ECMS adopters (Becker, Seidel, Pfeiffer and Janiesch, 2006, p. 3922). In this work, we propose a modeling language for mapping enterprise content called the ECML (Enterprise Content Modeling Language) that may serve as a starting point for later ECMS implementations.

The remainder of this article is structured as follows: According to a design science approach (Hevner et al., 2004; Simon, 1996), we first present the underlying design process and then analyze previous work in the field of ECM in order to derive common requirements on a modeling language for specifying enterprise content. Subsequently, we provide a theoretical foundation for our approach by presenting an established research framework for ECM. On the basis of both, the ECM research framework and the requirements that have been elaborated by consulting the relevant literature, we then introduce the ECML and provide a simplified application example. The application serves as illustration of concept and helps to identify future research needs. We finally conclude with a short summary and give an outlook on fields of further research.

DESIGN PROCESS OF ECML

The research methodology of this paper accords to the design science paradigm (Hevner et al., 2004; Simon, 1996) and therewith aims at the development of a modeling language—a so called “IT artifact”—which has to be both practically relevant and domain independent (Becker et al., 2006). The underlying design process of the modeling language is displayed in the following Figure 1. It is primarily based on the work of Rossi, Ramesh, Lyttinen and Tolvanen (2004), Greiffenberg (2004), Walsham (1995), Takeda, Veerkamp, Tomiyama and Yoshikawa (1990), Kamlah and Lorenzen (1984), and Becker et al. (2006).

![Figure 1. Methodology of the research process (cf. Becker et al., 2006)](image-url)

The awareness of the practical problem built up the initial point of the research process (phase I). It has primarily been identified during meetings and workshops with several medium- and large-scale enterprises being confronted with the challenge of efficiently adopting ECM. Within this context, the research question (“How to specify an organization’s individual content situation on the basis of a domain independent, conceptual modeling language?”) was also explicated. In particular it evolved that such a modeling language has to foster a sound communication between ECMS adopters and system users in order to identify and consider their specific information needs. Furthermore, the modeling language should hold a
degree of formality that serves as a basis for identifying and applying those ECMS functionalities best fitting the organizational and user-related requirements.

Design science comprises the definition of objectives which are consistent with prior research (Peffer, Tuunanen, Rothenberger and Chatterjee, 2008, p. 54). Accordingly, a comparative literature review has been conducted in process phase II in order to a) derive further requirements for a modeling language for enterprise content and b) identify available and problem relevant modeling languages/methods (cf. the section on related work). However, as we found out that these approaches do not fulfill all the relevant requirements, the decision was made to develop a modeling language from scratch.

Furthermore, design science aims at developing an IT artifact that draws from existing knowledge to ensure the rigor of research (Peffers et al., 2008, p. 49; Hevner et al., 2004, p. 83; Nunamaker, Chen and Purdin, 1991, p. 89). Rigor is thereby derived from an effective usage of existing foundations, e.g., theories and models (Hevner et al., 2004, p. 80). As we point out in the section on related work, the scientific interest in ECM has been very low so far. However, we identified an established research framework for ECM (phase III) that we transferred to the development of our modeling language (described in the section on the theoretical background). The ECML was finally constructed by first defining the required language constructs and then developing a meta-model on the basis of the Entity Relationship Modeling (ERM) method (phase IV, cf. the section on the ECML).

The hatched phase symbols in the figure particularly represent the current status of our research process and thereby also point out potential future work. Phase V is mainly inspired by the idea of cooperating with different organizations in each of the two iterations (as symbolized by the two arrows) in order to collect practical data for incorporating our findings and evaluating our modeling language. Within the first sub phase, case studies are set up on the basis of hypothesis referring to the following three general propositions:

- A conceptual modeling language is a requisite means for adopting ECMS (P1).
- An application of our modeling language supports the identification and application of required ECMS functionalities fitting the organizational, user, and content-related requirements best (P2).
- The conceptual specification of an organization’s content situation fosters communication between ECMS adopters and users and therewith provides a basis for later ECMS implementations (P3).

According to the next sub phase, suitable cases and adequate data collection techniques have to be selected. We hold that particularly document analyses may be applied for data collection (as documents already represent certain content assets). Furthermore, interviews and business process analyses come into question (as conceptual models of business processes usually contain information objects (content assets) and organizational units (content users/owners) as model elements (e.g., the EPC; cf. Scheer and Schneider, 2006). Within the second sub phase, experiences that have been gained during the application of our modeling language are collected in order to further elaborate the language within sub phase 3 (“Language Refinement”). The results that have been gained from the interviews and document/business process analyses are then compared to the propositions/hypotheses. Hence, the main objective of this phase is validating our language in order to adapt it where reasonable. After the adaptation—and if necessary—the language will be applied within additional organizations for further refinements (i.e., process V starts over again).

We are already applying the modeling language for specifying the content situation of a large-scale international enterprise which is currently implementing ECM processes worldwide. Therefore, a first case study has been set up. So far, several interviews and workshops have been conducted in different business areas possessing high relevance for ECMS implementation. The modeling experiences that have been gained within this case study have already been considered within the ECML definition presented in this paper. However, they are only implicitly discussed when introducing the language. Due to the confidentiality of the developed content model, we only present a simplified example that actually does not refer to the case (cf. the section on the exemplary application). However, we hold that this example serves as a suitable means to comprehensibly illustrate the applicability of our modeling language.

As the research methodology implicitly structures this paper, in the following section, we present selected results that have been gained during phase II (literature review).

**DERIVING REQUIREMENTS FROM A LITERATURE REVIEW**

Against the increasing importance of ECM, it has, up to now, mainly gained attention by practitioners. Especially technological aspects of ECM (particularly Web Content Management Systems (WCMS); cf. McKeever, 2003) have been picked up as central themes (cf. Browning and Lowndes, 2001). The scientific interest in ECM is in contrast still remarkably low (Tyrväinen et al., 2006, p. 627). We conducted a comparative literature review in the field of ECM and found out that
methodical support for adopting ECM can rarely be found in IS research literature (cf. vom Brocke et al., 2008). Within this context, particularly a suitable modeling language for conceptually specifying an organization’s individual content situation has not been developed yet. Though there are some “content-related” modeling approaches available (e. g., for e-learning content, cf. Thalheim, Binnemann-Zdanowicz and Tschedied, 2003), they all do not sufficiently fit to ECM.

Within this section, we present the main results from selected contributions that were part of the literature review. Thereby, in particular we consider the meaning and subject of the term “content”. The clarification of the content term serves as a basis for deriving requirements on our modeling language. An overview of contributions considered in the following is provided by Figure 2.

<table>
<thead>
<tr>
<th>Source</th>
<th>Main Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reimer, 2002</td>
<td>Structure and Functions of ECM Systems</td>
</tr>
<tr>
<td>Smith and McKeen, 2003</td>
<td>Development of an ECM Strategy</td>
</tr>
<tr>
<td>Rockley et al., 2003</td>
<td>Development of a Unified Content Strategy</td>
</tr>
<tr>
<td>O’Callaghan and Smits, 2005</td>
<td>Development of an ECM Strategy Framework</td>
</tr>
<tr>
<td>Nordheima and Päivärinta, 2006</td>
<td>Identification of Critical Issues in ECM Systems Development</td>
</tr>
<tr>
<td>Päivärinta and Munkvold, 2005</td>
<td>Development of a Framework for ECM Implementation</td>
</tr>
<tr>
<td>vom Brocke et al., 2008</td>
<td>Development of a Framework for ECM Analysis and Specification</td>
</tr>
</tbody>
</table>

Figure 2. Selected contributions in the field of ECM

Taking a mainly practical oriented perspective on ECM, Reimer (2002) primarily refers to ECMS functionalities and structures. Therewith, the term “content” is only considered to a small degree. From his point of view, once information is captured it becomes content (p. 18). As a result, content can only exist in a digital form (p. 17). Examples for content are computer generated reports, photographs, audio, and video (p. 17). Therewith, Reimer solely differentiates content on the basis of its format. This quite limited perspective on content is emphasized by stating that content assets are often shared as e-mail attachments (p. 18). However—and as we point out in the following—an e-mail itself can be seen as a batch of different content assets (e. g., text, address, signature, date). Hence, Reimer’s approach falls significantly short of the content term.

Smith and McKeen (2003), O’Callaghan and Smits (2005), and Rockley, Kostur, and Manning (2003) focus on the development of an overall ECM strategy. Smith and McKeen state that ECM first of all has to consider what content has to be included (p. 648) and further hold that there are different “types” (p. 649) and “forms” (p. 651) of content. However, they do not sufficiently point out the differences between a “content type” and a “content form.” Accordingly, O’Callaghan and Smits explicitly underpin that ECM involves the analysis of content needs and decisions on which content has to be managed and how it will be managed (p. 1271). As well as Rockley et al., O’Callaghan and Smits do not understand content assets as complete documents, like reports or marketing brochures, for example. In fact, they emphasize that ECM is not about managing entire documents, but re-using content to be assembled in a multitude of documents; e. g., the same content asset may be used in a brochure and in a press release at the same time (Rockley et al., 2003, p. 23ff.). Furthermore, they state that content is characterized by a certain structure (O’Callaghan and Smits, 2005, p. 1273).

The report from Nordheim and Päivärinta (2006) deals with an ECM case study that serves the purpose of deriving key issues and challenges for general ECM implementation. They particularly emphasize that the management of content has to regard content lifecycles as well. A content lifecycle consists of content producing, capturing, versioning, distributing, publishing, retrieving, and archiving (p. 649). Päivärinta and Munkvold (2005) present an overall content model for ECM that contains structure, view, layout, lifecycles, metadata, and corporate taxonomy (p. 3ff.). Beneath the elements that have already been explained, the corporate taxonomy refers to a logical and conceptual structure providing the basis for searching for content, e. g., by automatically defining metadata. In contrast, the layout refers to the presentation of content. Finally, within our prior research, we already presented an overall framework for ECMS analysis and specification (vom Brocke et al., 2008). It is primarily based on the definition of content properties (e. g., the degree of a content asset’s structure). Within this work, we particularly pointed out the challenge of assessing an organization’s individual content situation and the demand for methodical support.

In conclusion, the presented articles provide different content properties that we used to derive the following requirements on a modeling language for enterprise content:
A conceptual modeling language for ECM has to consider different content needs and therewith to differentiate specific user rights (R1). For example, one user might not be allowed to edit but only to view certain content assets, whereas another one even possesses the rights for deleting them.

Content assets are characterized by specific content lifecycles. As ECMS provide functionalities particularly referring to these lifecycles, they have to be specified in order to identify content-related functional requirements (R2). For example, an employee record usually possesses a high degree of confidentiality. Hence, it does not pass an external publishing phase like web content, for example. Therewith, it further does not require related WCMS functionalities.

An important aspect of ECM is the separation of content, structure, and presentation. Hence, a conceptual modeling language has to provide means to align presentation forms and predefined structures to different content assets (R3).

Depending on the modality of ECM, an electronic document can be seen as a batch of different content assets or as a content asset itself. Hence, a conceptual specification of such content assets and containers has to consider in particular the relationships between them (R4).

In addition to the derivation of these requirements, our literature review also showed that they are not sufficiently fulfilled by existing modeling languages/methods from related IS disciplines (e.g., the Multidimensional ERM; cf. Chen, 1976). However, due to word limitations, we will not go into detail on this (cf. Becker et al., 2006, p. 3926f. for more detailed reasoning). In the following, these requirements are used in order to develop our modeling language. Afore, a theoretical foundation for our approach is given by presenting a research framework for ECM.

THEORETICAL BACKGROUND OF ECML

In accordance with the findings presented in the previous section, Tyrväinen et al. (2006) state that ECM received far too little attention within the related research communities. They present a general framework for ECM research that served as a foundation for developing our modeling language. The framework mainly contains four research perspectives that need to be addressed when adopting ECM: enterprise, processes, content, and technology (cf. Tyrväinen et al., 2006 and Figure 3 in the following).

![Figure 3. A Research Framework for ECM](image-url)

Research questions on the enterprise level refer to institutional, especially social and legal aspects (e.g., compliance, archiving times). The process view refers to both, process development and deployment. However, we hold that a conceptual specification of an organization’s individual content situation particularly has to take the content and the technology perspective into account. This can be reasoned by proposition (P2) (a conceptual model “serves as a basis for identifying and applying those ECMS functionalities fitting the […] content-related requirements best”, cf. the section on the design process). The content perspective therefore comprises information about content (e.g., structure and presentation) and users (e.g., user rights). As the main objective of specifying content is identifying and applying the most suitable ECM technologies, we
arranged the technology perspective as “outcome” at the bottom of Figure 3. It comprises ECM hardware, software, and standards that are applied for managing the content.

Recapitulatory, our approach is mainly inspired by the idea of specifying content referring to content users and information (content view) in order to identify those ECMS functionalities (technology view) that fit the organizational requirements (enterprise and process view) best. On the basis of these findings, in the following section, we present the Enterprise Content Modeling Language (ECML) by referring to both, the requirements that have been derived within the literature review and the ECM research framework.

THE ECML – A MODELING LANGUAGE FOR ENTERPRISE CONTENT

As a design technique we apply meta-modeling. Meta-modeling is a widely established means for specifying conceptual modeling languages, widely making use of the Entity Relationship Modeling (ERM) method (Becker et al., 2006). Accordingly, the ECML definition has been elaborated on the basis of the ERM displayed in Figure 4. In this paper, we instead refer to the graphical representation of these element types which is given by Figure 5 on the next page. Within this figure, the ECML element types are arranged referring to the introduced ECM research framework on the basis of the technology and the content dimension (accordingly, the latter is separated into information and users). For the sake of a preferably clear presentation of the ECML, modeling steps are symbolized by round brackets in the follow-up (please also note that italic font style is used when referring to an element type).

![Figure 4. ECML entity relationship meta-model](image)

Referring to the technology view, the ECML enables one to combine available ECMS Functionalities to ECMS Functionality Groupings. As ECMS are most often structured modularly (Böhn, 2007), these groupings (may) also represent certain ECMS modules. The definition of ECMS Functionality Groupings serves as a basis for specifying Content Lifecycles (1). As introduced in the section on related work, a content lifecycle comprises certain Lifecycle Phases that have to be supported by specific ECMS Functionalities (or ECMS Functionality Groupings). Hence, the second requirement of an ECM modeling language (R2) is fulfilled by the ECML. Referring to the content view of the ECM research framework, content users can be defined on the basis of the same named element type (2). To further enable the specification of user rights—and in order to therewith meet the requirement (R1)—different Content Lifecycles (representing ECMS Functionalities) are assigned to the Content Users who can finally be combined to User Groupings (3).

The information perspective of the ECM research framework is considered by the three ECML contexts Content Properties, Content Assets, and Content Containers. According to requirement (R3), the ECML enables one to differ between certain Content Structures and Presentations (4). On the basis of all the element types being explained so far, certain Content Assets can then be specified (5). Accordingly, a Content Asset comprises a Content Structure and Presentation, a Content Owner (specified as “regular” Content User, cf. the ECML meta-model), a User Grouping, one (or more) ECMS Functionality Grouping(s) (representing all the required ECMS functionalities referring to the Content Asset), and the Content itself (however, a Content Asset actually must not possess Content, e. g., an organization’s logo). According to requirement (R4),
Content Assets can finally be combined to Content Containers (6) that of course may also possess their own Structure or Presentation, for example. Accordingly, within the meta-model, we used the general entity type “content aggregate” to differentiate between a Content Asset and a Content Container.

As symbolized within Figure 5, the ECML also differs between element definitions and occurrences (the latter are typed in italic font style). For example, ECMS Functionalities are defined in the context of the same name (element definition) and are then re-used in the Content Lifecycle context (element occurrence). This concept of re-usage makes it possible to update all occurrences of an element by only editing it one-time within the context of its definition.
EXEMPLARY APPLICATION

In this section, an exemplary application of the ECML is given in order to provide a deeper insight into its applicability. The content model has been developed with the (meta-)modeling software H2-Toolset (cf. Delfmann, Janiesch, Knackstedt, Rieke and Seidel, 2006). A screenshot of the developed content model is given by Figure 6. For illustration purpose, from now on, we refer to model elements contained within the figure by using italic font style.

ECML element types are illustrated on the basis of the paper at hand: The AMCIS Paper: “Towards the Specification of Digital Content – The Enterprise Content Modeling Language (ECML)” was collaboratively written by four authors. Hence, especially the usage of ECMS functionalities supporting the process of Collaboration (e. g., Versioning, Access Protection, or Archiving) would have to be considered when applying an ECMS. Furthermore, the paper generally belongs to Conference Proceedings that are usually Created, Edited, and Deleted by the authors (but not captured, for example). In terms of Collaboration, these Content Lifecycle Phases require support by different ECMS Functionalities (e. g., Access Protection is not required when Creating a paper). Referring to Conference Proceedings, the four Content Users possess different User Rights, e. g., one might not be allowed to Delete but only to Edit a conference paper. As the authors are working at different IS Departments, furthermore, they are grouped. Within a more general ECM context, user groupings are of special importance in case different organizations are working together in terms of a so called “Cross-ECM”.

Figure 6. Screenshot of an exemplary ECML model

On the basis of different content presentations (conference Templates) and content structures (e. g., Case Study, Research Paper), all the relevant modeling elements are then combined to the conceptual specification of the paper. Please note that the
occurrence of the ECMS functionality grouping collaboration does not comprise the archiving function at this point (though it is contained within the original definition). Hence, an author who is generally allowed to apply this function referring to Conference Proceedings (e.g., the first author) is not able to use it for this paper (as this is generally not designated). Finally, the paper might be used by one of the authors within her/his cumulative PhD thesis. In that case, presentation and/or structure can also be adapted within the context of the content container PhD Thesis.

This example may illustrate how to make use of the ECML in order to capture the various facets of content within an application area.

CONCLUSION

With this paper, we proposed a conceptual modeling language for specifying an organization’s individual content situation, the Enterprise Content Modeling Language (ECML). Therewith, we intend to contribute to an emerging field of IS research: Enterprise Content Management (ECM). According to a design science approach, we first presented related work in the field of ECM in order to derive specific requirements on a modeling language for ECM. We then considered an established research framework for ECM serving as theoretical foundation for our approach. Referring to both, the research framework and the requirements that have been identified by consulting the relevant literature, we then introduced the definition of the ECML on the basis of an entity relationship meta-model and presented a simplified application example.

Our work on ECM is essentially driven by the perception that there are various types of content relevant to making use of the various ECMS functionalities when designing ECM aware business processes. Therefore, the analysis and specification of content types is crucial for the implementation of ECM systems (ECMS). The ECML may provide a basis for this. We particularly hold that the ECML may foster a sound communication between ECMS adopters and system users in order to a) identify their information needs and b) apply those ECMS functionalities fitting to the user requirements best. According to the presented design process of the ECML, future work will focus in particular on further refining the modeling language and its application within additional industry projects.

REFERENCES