A Reference Model for Flexible Content Development

Sami Jokela
sami.jokela@hut.fi
Helsinki University of Technology
P.O.B. 9555
FIN-02015 HUT
Finland

Janne Saarela
jsaarela@w3.org
World Wide Web Consortium

Abstract

Media industry is changing fast. Instead of spending years on product development and testing, media companies today have to develop complex new products for a variety of media platforms in a matter of days.

Our work with the media companies has shown that unless the production process is understood, rethought, and altered, only fraction of the possible advantages of separating content and formatting will materialize.

This paper introduces a reference model for multi-purpose publishing by dividing the electronic publishing process into components and process steps. The relation and interaction of these elements is also discussed in detail. We believe that the framework will assist in gaining better control of the existing and planned electronic publishing processes and will lead to better re-usability of the content on multiple products and platforms.

1. Introduction

Electronic publishing is facing new challenges. Content providers must produce and manage content for an increasing variety of media platforms and customer groups. In practice, the publisher has to publish the same content on platforms that vary in characteristics such as portability or available functionality. These challenges call for better content management and re-use.

The goal of re-using content on multiple products and platforms affects the entire life-cycle of electronic publications. Authoring, publishing, and delivery require careful planning in order to harness the advantages of reusability.

These challenges create new opportunities for publishers. Whereas one publisher focuses on authoring original content in video format, another may only focus on packaging content originating from multiple sources to be delivered on CD-ROMs.

In this paper we develop a reference model for multi-product multi-platform publishing by analyzing the components and steps of the electronic publishing process. We also study the dependencies between different process steps. Based on the results we then suggest optimization possibilities of the overall process.

A number of books and articles have relevance to our work. For example, [9] and [7] both contain multiple articles related to semantics, metadata, and managing multimedia content. These sources are especially useful when planning the technological aspects of the electronic publishing.

[11] discusses different aspects of achieving higher efficiency in the development of complex products. Although the main emphasis of their article is in managing design data, the idea of observing design processes from different dimensions i.e. perspectives has clearly affected our work.

[3] models a general publishing process where the process steps are submission, acquisition, quality control, production, and delivery. These steps assist in building an automated publishing service. However, this paper does not address the problem of multi-platform support nor recognize additional process steps such as personalization or automated layout we consider essential for electronic publishing.

2. Four layers of electronic publishing

It is easier to understand and manage electronic publishing if we identify and understand its components and structure. For this purpose we present an abstract framework that consists of four different layers [Figure 1]. We believe that these four layers are able to express the complexity of publishing the same content in multiple products on multiple platforms.

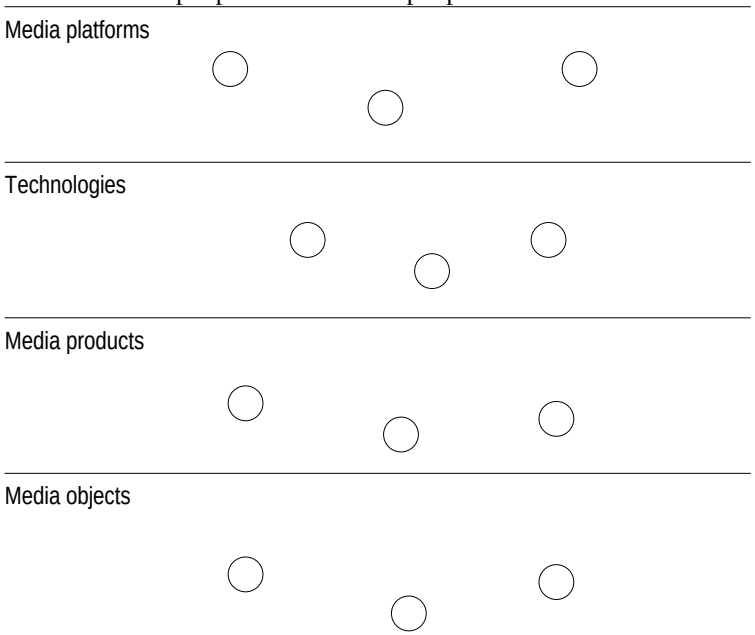

Figure 1: Four-layer framework of electronic publishing

The first layer is media platforms, on which the content is delivered and presented to the user. This layer contains terminals and other possible platforms such as paper. The second layer is technologies. Technology layer acts as an enabler and implements the functionality of the media products on the media platform. In many cases the technologies are built-in into the media platform. However, if the technology is not identified and treated separately, we may fail in optimizing the electronic publishing process on multiple media platforms. The third level is media products, that form an identity for a group of media objects. The last layer is media objects, that consist of both the actual content and descriptions of content qualities. The following sub-sections discuss these layers in detail.

2.1. Media platform

Media platform is the medium that is used to consume the content. In most cases media platform is equivalent to the end-user terminal. In its simplest form consuming content means presenting, i.e. reading, listing, or watching it. If the medium and the content support more sophisticated functionality, consuming may also involve operations such as personalization or interaction.

If the same content is to be used on multiple media without reproducing it for each media separately, one must understand the nature and possible capabilities of each media platform. The following table presents a number of key capabilities of some possible media platform alternatives. The values in the table can be in some cases argued, but the most important issue with the table is to understand, that different media platforms have different capabilities.

Capability	A computer with Internet access	A computer with local mass media, e.g. CD-ROM, DVD	Radio	Mobile phone (GSM)	Digital TV	Paper
Network-dependence	Yes	No	Yes	Yes	Yes	No
Interactivity	Yes	Yes	No	Yes	Yes	No
Personalization	Yes	No	No	Yes	Yes	No
Sense of interacting communities	Yes	No	No	Yes	Yes	No
Real-time audio	Yes (where possible)	Yes	Yes	Yes	Yes	No
Real-time video	Yes (where possible)	Yes	No	No	Yes	No
Multicast delivery	Yes (where possible)	Yes (Shipping to customers)	Yes	Yes	Yes	Yes
Updateable content	Yes	No	Yes	Yes	Yes	No
Location awareness	No	No	No	Yes	No	No
Price to consume content (access)	Low to medium	Low	Low	Low-High	Low-High	Medium
Price to produce content	All categories	High	High	Low	High	High
Portability	No/Yes	Yes	Yes	Yes	No	Yes

Table 1: Characteristics of different media

2.2. Technology platform

Each media platform contains a palette of tools i.e. technologies that are available to produce and to consume the content. This set of technologies is called the technology platform. On some media platforms the technology platform is fixed, as is the case with radio receivers. With other media the technology platform is expandable for example by downloading a new plug-in for the Web browser. The technology platform may also contain alternative methods to implement a certain capability. For example, a Web publisher may implement interactivity using forms in the hypertext language HTML or alternatively using Java programming language.

Figure 2 illustrates with a simplification the relations between a medium, technologies, and capabilities. In the example User input can be implemented either with voice recognition or by entering text. For the capability Presenting pictures there are two possibilities, GIF or JPEG image formats. If we want to have both the capabilities on one medium like Internet, we need to use Text - based communication, but can select either GIF or JPEG for Presenting pictures.

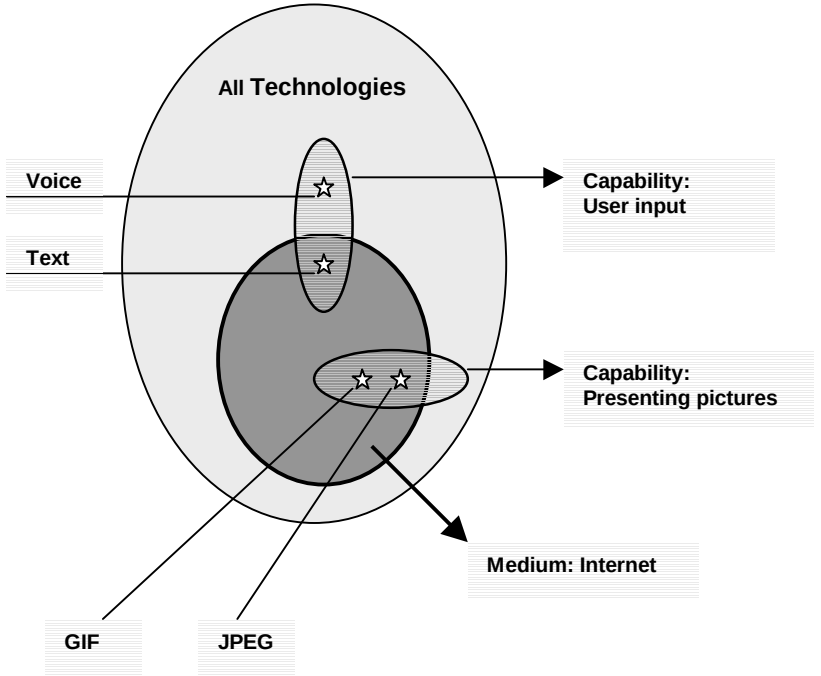

Figure 2: Definition of the technology platform

2.3. Media products

Media products are in many ways equivalent to physical products. Media products package content according to some predefined rules. For example, a web magazine is a media product that may contain a number of separately formatted stories.

One goal with media products is to keep technology platform, media products, and content composition separate [Figure 3]. By doing so, media products can be defined as platform-independent templates containing many media objects.

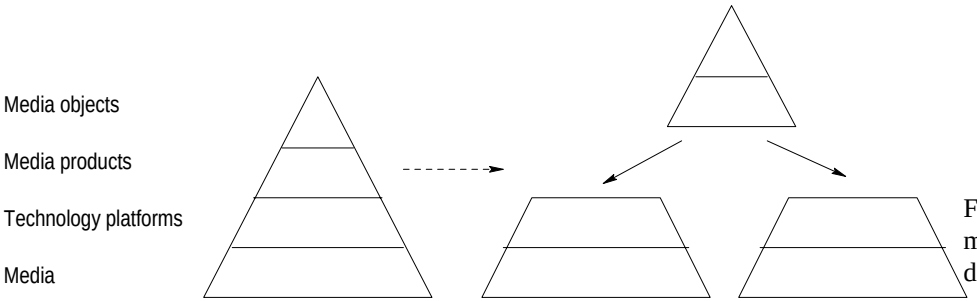

Figure 3: Separating different layers of media products

2.4. Media objects

Media objects are the basic building blocks in electronic publishing. These objects contain both the actual content and a description of it. Generally, these descriptions are called metadata i.e. information about information [10]. This metadata describes the characteristics of a given media object in a machine-understandable form. For example, metadata may describe the format or semantics of the content.

2.5. Justification for the content decomposition

We believe the four-layered decomposition of content is valid because of the following reasons.

- The complexity of publishing can be better understood — Whereas a publishing process may look trivial for a single media product to be delivered on a single medium, the relations between multiple media products and multiple media may be easier to understand with the decomposition.
- More layers increase the complexity of the model — Adding more layers to the model may complicate the understanding of the key issues in multi product multi platform publishing. For example, if we add customers as a fifth layer in the model, that undermines the inherent existence of customer aspects on each layer of the model.
- Model simplification results in loss of valuable information— For example, if we remove the technology platform from the model, we could describe the re-use of media objects and media products. However, we would then be missing a very important complexity factor and not be able to understand and optimize the electronic publishing process properly.

Figure 4 presents an example, where multiple media products share common media objects. These media products use different technology platforms to deliver the content to multiple media.

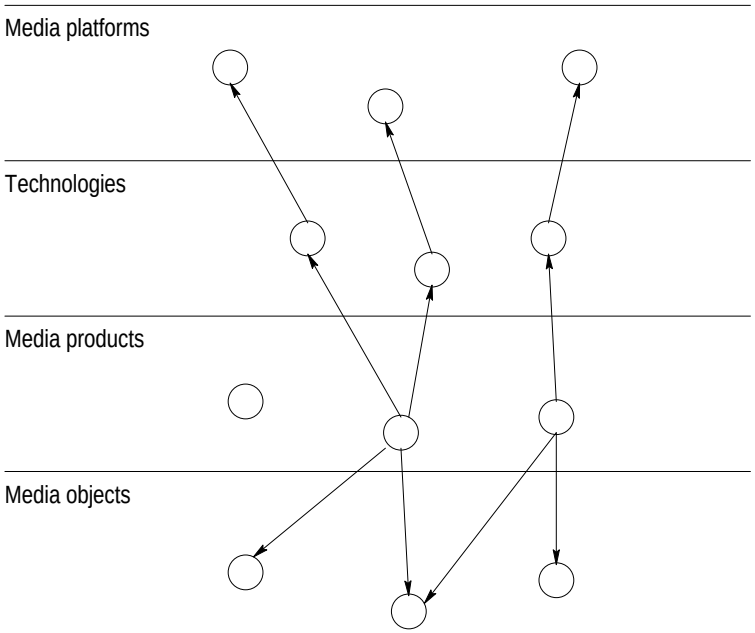

Figure 4: An exemplary publishing scenario

3. Process decomposition

In the previous discussion we have established a four-layered model of electronic publishing. This model helps to understand different components involved in the electronic publishing process. Next step is to determine what kind of effort the multi-product and multi-purpose publishing requires. For this purpose, we present a process decomposition that describes the connections between the elements in the four-layered model.

We decompose the process to implementation independent steps. This implies information about *what* the step is supposed to do instead of *how* it should be performed.

At the highest level we divide the process into three abstract process steps: Authoring, Content composition, and Delivery [Figure 5]. These are presented in the following sub-sections.

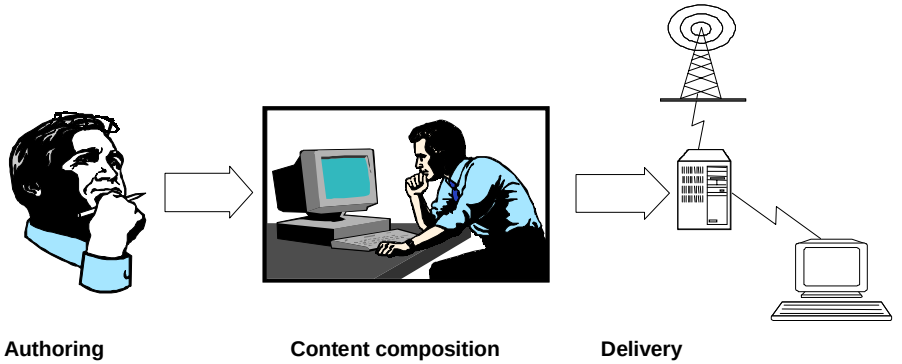

Figure 5: High-level steps of content development

In addition to the three high-level process steps, product development activity runs in parallel with the content development. Product development is elaborated later in this chapter.

3.1. Authoring

Authoring is the first step in the framework. Authoring consists of a sequence of steps that involve creating or acquiring the actual content and encoding that data in a computer understandable form. Once the authoring is completed, we assume that the computer can automatically manipulate the content, e.g. by selecting suitable content to be published in an online newspaper.

We exclude the layout design and product-specific layout work from authoring, although in some cases these operations are performed together.

The authoring may vary within the following two dimensions:

- Time dimension - Publishing may be performed as a batch work initiated by a deadline set by, for example, a publishing schedule. Publishing may also occur in real-time as is the case with live broadcasts.
- Platform dimension – Authoring may result in content that can only be used for single purpose on a specific medium platform. Or, in an ideal case the authoring can be accomplished in a purely platform-independent fashion. The separation of content from its presentation enables the content reusability on multiple platforms. For example, text can be authored using

structured markup document formats such as SGML [6] and XML [2] that are inherently independent of any platform.

Once the authoring step of an electronic publishing process is completed, the produced media objects are available for the next process step, content composition.

3.2. Content composition

Our next process step is content composition. Content composition uses media objects and produces media products that are ready for delivery.

In addition to media objects, content composition requires information on media product structures. This information expresses the properties of media products, namely a specification how the media objects are presented and what functionality is used in the media products. In addition, this process step may use customer feedback. This feedback — either direct or indirect — allows the content composition to take into account personal preferences. This enables personalization of content and presentation.

Again, the result of this process step can be either platform dependent or independent. If the media product adapts its content to user interests, that personalization can be performed platform-independently. For example, the user can define only to receive news headlines about sports. These headlines can be selected in the content composition step to be further delivered to the user and finally presented given the user's personal presentation specification.

Personalizing the presentation of a media product typically leads to defining platform-dependent qualities. For example, a user may request news headlines to be presented in a temporal presentation without the need for user interaction. This would involve encoding the presentation with a specific technology such as Synchronized Multimedia Integration Language [4] or [5].

If the media product is produced only once, there is no need to formalize the structure of the product into a reusable product template. If the media product is published multiple times, it is advisable to use templates to allow re-use and automation during the publishing. These templates are generated and maintained in the product development process.

Content composition with templates requires a varying level of work depending on the case. In the simplest situation, templates can be directly applied assuming that a given set of media objects is authored and available. Next, the content composition process step can be automated by attaching the media objects to appropriate slots in the media product specification. For example, a daily newspaper can have the same media product templates from day to day with the individual stories i.e. media objects changing for each issue.

A more complex situation occurs when the media product template requires changes. For example, the structure of an online newspaper is changed to include a new section on domestic politics. This change requires an engineering effort to update the product template. In addition, the authoring process must result in media objects that can be associated with the new section.

The most complex situation occurs when the product template requires modifications in the technology platform. This means that the needed media product functionality cannot be implemented using the existing technology platform. For example, interactivity is added to the Web newspaper by enabling users' annotations on the individual articles. To support this new feature the publisher has to build both the production system that supports interactive newspaper as well as to ensure that the customers have browsers that support this new functionality.

3.3. Content delivery

Our last high-level process step is content delivery. This step begins after the content composition and extends to the point in time when the customer has access to the media product.

We define content delivery as the distribution of the media product to the customer. The customer may not be the final end-user but simply a next step in the distribution chain. Also, if the customer retrieves content directly from the publisher, the publisher does not effectively deliver anything. In that sense the content delivery process step contains both push- and pull types of delivery.

Content delivery can also be personalized. This implies the customer can change the preferences of the content delivery process step. For example, the

preferences may express the type of media products to be delivered as well as the time when the delivery should take place.

3.4. Product development

Product development contains those steps that are required to define and maintain different components that are used during electronic publishing. These components include media object templates, media product templates, and the technology that is required to enable the functionality of media products. In addition, product development ensures that there exist tools and methods for producing media products during the electronic publishing process. This structure is described in *Figure 6*.

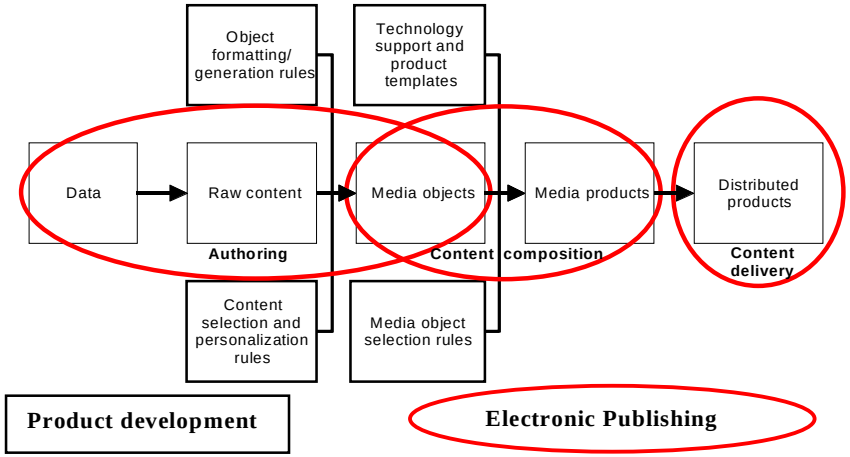

Figure 6: Components of the electronic publishing process

Even though product development has steps similar to electronic publishing, these two processes have fundamental differences. For example, in most cases the templates must be available before any electronic publishing takes place. Template creation and modification usually takes place periodically and requires project-like effort, whereas electronic publishing resembles normal production and is an on-going effort. The processes are also different in terms of required resources. Templates typically require media product, design, and technological skills, whereas electronic publishing concentrates more on content expertise and creativity. Although these differences exist, successful electronic publishing requires close interaction between these processes. In this way we can ensure that electronic publishing process stays in optimal shape.

3.5. Pre- and post-conditions of process steps

This section will decompose the three high-level process steps we introduced earlier: authoring, content composition, and delivery. We try to give an idea of the potential steps an electronic publishing process may use. Our decomposition is not the only possible one, but reflects the steps and detail level we consider worth including in this framework.

We will assign a unique identifier to each process step, e.g. Ai means platform independent authoring. This identifier will help us present process chains i.e. sequences of steps in a more compact way. In addition, we will present pre- and post-conditions of each step. The pre-condition defines the information that is required to perform the process step. The post-condition defines the output from the step. These conditions will help us analyze dependencies between the steps. We will later use these dependencies to analyze whether re-ordering the steps or running some of them in parallel will help to optimize the process chain.

3.5.1. Authoring

Ai Content is authored in platform-independent format. For example, raw textual material is authored without layout design for a certain page size.
Pre-conditions: none
Post-conditions: Platform-independent media objects

Ad Content is authored in a platform-dependent format. If the same format can be used on several media, this process step is partially platform-independent and can be represented as Ai. For example, if a video stream is encoded in some specialized streaming format, this task is denoted by Ad if it can be used only on one particular medium. If the streaming format is supported on multiple media, it is denoted as Ai. Typically, given a rich presentation environment of a medium platform, the authoring may have to involve integration of specific stylistic

features and interactivity. This type of authoring may make the media objects very dependent on the medium.
Pre-conditions: none
Post-conditions: Platform-dependent media objects

Ap Content is authored in platform-dependent format for a one single media product, which is the strictest form of authoring. For example a one-time-only multimedia presentation may contain such content.
Pre-conditions: none
Post-conditions: Platform and media product dependent media objects.

3.5.2. Content composition

Sc Selection of the media objects to be included in a media product from the media objects available. This operation has varying levels of dependence. Selection means simply selecting a subset of all media objects. Media objects can be selected also based on user preferences.
Pre-conditions: media objects
Post-conditions: subset of media objects

Sp Selection of the layout template.
Pre-conditions: none
Post-conditions: layout template for the media product

Ld Layout formatting associates media objects with presentation methods provided by the technology platform of the target media. This design may involve only generic presentation features in which case the resulting content is still usable on multiple technology platforms. However, depending on the amount of features used, the result may be usable on one technology platform only. For example, adding stylistics features to text may still keep the text usable on multiple platforms whereas setting parameters to speech synthesizer for generating vocal output limits drastically the possible technological alternatives.
Pre-conditions: available content
Post-conditions: platform-dependent layout

Lp Product dependent layout formatting. This is the same than Ld with the exception that this step results in a unique media object layout for the final media product. An example of this is a generation of a front page for an online newspaper.
Pre-conditions: available content
Post-conditions: platform and media product dependent layout

Pc Personalization of content. Content is modified according to some rules. For example, content is personalized for a certain customer by adding customer name to the owner field of the content.
Pre-conditions: availability of authored content, instructions on how to perform the personalization.
Post-conditions: personalized content

Pf Personalization of the supported functionality. For example, the customer can control the presentation of individual media objects by so-called style sheets.
Pre-conditions: functionality is implemented and personalization instructions exist.
Post-conditions: personalized implementation of functionality

P Publication of a media product. During this step a deliverable version of the media product is created. A media product template is applied to the media objects to create a completed and media dependent product. If the customer does the final formatting, this process step is not performed by the publisher. Should the media product have physical components, they are produced in this process step. For example, one can produce a master and distribution CDs, or print the pages on paper. With real-time content this means the production of the resulting real-time stream.
Pre-conditions: completed and available media objects, media product template (if used), and implementations of the functionality
Post-conditions: Completed platform-dependent media product

3.5.3. Content delivery

Dp The personalization of the delivery that takes into account an individual user’s preferences. For example, a user may have asked a Web online newspaper to send the contents to her/him through email at a certain time. If the delivery personalization criterion is not changed frequently, this process step can also be part of the product development process.
Pre-conditions: completed platform-dependent media product, user preferences or other triggering events for delivery
Post-conditions: Media product ready for delivery

D The delivery of a media dependent media product. Typically every media product requires a delivery mechanism that is largely dependent on the presentation media. For example, a Web online newspaper may send its files using an Internet file transfer protocol (FTP).

Pre-conditions: Media product ready for delivery
Post-conditions: Delivered media product

3.5.4. Product development

Cf Implementation of the media product functionality using different methods and technologies. Implementation ensures that both the production system and the target media can support the required functionality. For example, a voting system requires a media product development system that allows the publisher to use voting in the media products. In addition, the media platform must support voting functionality, e.g. interaction, when the content is consumed. If production or distribution cannot support the functionality with existing methods and technologies, they must be developed.
Pre-conditions: none
Post-conditions: functionality implementation

Cp Creation of the media product template. Definition of the rules that are used to select the content and layout for the media product. These rules typically refer to the characteristics of the media objects such as their topic or format.
Pre-conditions: none
Post-conditions: media product template

Cc Creation of the media object template. This typically assists in generating individual media objects. By using such a media object template, the author of the media object can be assisted by automatically adding, for example, metadata or signature to the content.
Pre-conditions: none
Post-conditions: media object template

3.6. Re-arranging process steps

The previous subsection has presented steps of an electronic publishing process. Given the pre-and post-conditions of each step we are now able to analyze the dependencies of these steps. The following list presents some restrictions on the process steps sequences.

Our simple notation to describe process dependencies contains two operators, < and =. < implies that a step occurs before another step. = implies that two steps are interchangeable.

- A = Sc — Content is either authored or selected from externally produced content
- A < D — All kinds of authoring must precede delivery
- A < P — All kinds of authoring must precede publishing
- L < D — Layout formatting must precede delivery
- L < P — Layout formatting must precede publishing
- A < Pc — All kinds of authoring must precede content personalization
- Pc < D — Personalization of content must precede delivery
- Pf < D — Personalization of functionality must precede delivery
- Pc < P — Personalization of content must precede publishing
- Pf < P — Personalization of functionality must precede publishing
- P < D — Publishing must precede delivery
- C < * — Template creation (functionality, media product, media object) must be the very first step if templates are used.

Using the previous list we can generate different processes that consist of different steps. Our purpose is not to produce an exhaustive enumeration of all sequences but instead to present via an example how the model could be used in real-life cases.

Example 1. A media company produces a series of encyclopedia on CD-ROMs. The media objects i.e. the individual articles contain multimedia content such as structured documents, images, audio, and video streams that are originally encoded in multiple formats. The media product is defined as a set of criteria such as people or locations with respect to the available metadata. The technology platform is in the control of the publisher by using stand-alone software that is shipped with the content on the CD-ROM. The final medium platform is targeted for home PC users that can run the software on the CD-ROM.

Naïve process sequence that implements single product on the given platform would become:

Cc Cp Cf Ad Ld P D, which means that after the templates for media objects (Cc) and for the CD-ROM media product (Cp) have been created and required functionality implemented (Cf), media objects for the medium are authored (Ad) and layout formatted (Ld). When media objects are ready, the CD-ROM is compiled, CD-ROM master is created (P), and distribution CD-ROMs are delivered (D) to the customers.

Now, the goal to publish the same product on the Web introduces the following changes:

A new product development activity is required to implement capabilities on the new technology platform, which means that additional Cf and Cp steps need to be added. The media object level requires no changes since the content object format can be — with the help of automated conversions — re-used on the new technology platform.

New process that implements the product on two platforms would become:

Cf1 Cf2 Cc Cp1 Cp2 Ad Ld P D

It is worth noting that the development of the Cf1, Cf2, Cc, Cp1, and Cp2 could be performed by a separate product development team. Once finished, the results can be used in the continuous production consisting of the steps Ad, Ld, P, and D.

4. Process optimization

We have previously introduced the components and structure of the electronic publishing process. This chapter builds on that information and presents results on how to optimize the process.

As a general rule the less manual work is needed during the electronic publishing process, the faster and cheaper the production is in the long run. Machines cannot substitute humans in the creative work, but computers can assist in minimizing routine work such as conversions between formats.

Equally important is adding re-usability and flexibility to the electronic publishing process. If the content is re-used on multiple products and platforms, the re-usability and flexibility become even more important. The following list describes some issues that have a major impact on the flexibility and the need for manual work.

1. Format issue — Are the media objects directly re-usable or do they require manual tailoring for each delivery platform? Do we need additional work to convert raw content into usable form? Do we produce media objects in a format that requires least manual work further in the process?

2. Content structure issue — How are the media objects combined for a given media product? Is each media product structurally unique or can we encapsulate their structure in a re-usable product template? Does our media product use the optimal set of technologies to implement its functionality? How can the media objects be addressed separately for more individual services such as searching or interactive publishing?

3. Production process issue — How can the progress of the production be tracked especially when there are several delivery platforms to be supported?

4. Product development issue — How much manual work and control metadata is needed to create new media products or to alter existing ones?

We discuss these issues in detail in the following subsections.

4.1. Format

Media objects should be as modular and flexible as possible. The ultimate goal is that each media object is directly usable for any media product on any media platform. In reality this is often difficult to achieve and may complicate the content creation too much. The pragmatic approach is to develop automated tools to manipulate the media objects to adapt to the delivery platform with as little manual work as possible.

The flexibility of the content objects can be guaranteed by using high quality encoding of content such as raw audio, vector graphics, pixel images, and

structured documents. These formats allow greater variety of conversions to different final presentations. It is therefore advisable to aim at preserving the source content quality as high as possible in order to allow easier convertions to final media products. For example, with structured documents the high quality implies high level of abstraction in the markup.

One method to optimize the process is to present the media object encoding characteristics in a separate metadata description. This description can be used in producing the media product without the need to access the actual content until publication. The metadata description may also exist without the actual content object. In this case, it signals a missing content object planned for delivery.

Media objects should always be associated with their metadata descriptions. If the only information available about a media object is its name, we cannot ensure that the object is properly used further in the electronic publishing process.

4.2. Content structure

Electronic publishing should be guided by using product templates that contain instructions how media objects are selected and used in the product. To allow automation in that process those templates should contain abstract composite relations in a machine-usable format. Minimum requirement for using templates is to define each media object with a unique identifier within the production environment. This allows machines to automatically store relations between the media objects such as version information.

Unique identifiers of media objects and products should be preserved from production run to another. Unique identifiers can be issued either centrally, or in a distributed environment each separate location can issue its own unique address space associated with the location resulting in a unique global identifier.

The challenge is to use applications that are aware of this unique addressing. We can, for example, require every media object modification to be stored in their metadata entries and that old versions are never overwritten or deleted.

In addition to unique addressing, each media object and product should have status information attached to it. This information can then be used to define dependencies between different steps of the production process. If certain status of the product has not yet been achieved, the process cannot continue to the next stage.

4.3. Production process

If we implement the previously presented recommendations for media objects and media products, the production process can be monitored and controlled. When the composite structures of media products contain certain properties such as format or size, we can query their state. A query can tell us whether a media product satisfies the given conditions and if not, what parts are non-valid. Running such queries frequently assists in understanding the production status at any given time. It is worth noting that such tools do not need to access the media objects at all, just their metadata descriptions.

4.4. Media product

Media objects cannot be used without bundling them into media products. Media products in turn use different technologies to realize their functionality. When the media product is produced, these technologies are used in the production and/or are added to the final media product. It is important to keep the development of the new technology and products separate from the actual production of media products.

The basic structure of the media products must be planned beforehand and characteristics of the resulting products must be understood. Required functionality must also be developed before the actual production. The important factor here is the characteristics of the customers. If their requirements change frequently, the product structures must be easy to alter and modify.

The medium of the product sets some constraints on its implementation and functionality. If the medium requires physical components in the delivery such as paper, content production and delivery is much more costly than with purely virtual products. Therefore we recommend that as much as possible of the product should be virtualized. A good example of this is a newspaper. If

personalization is fully exploited, each paper will be printed separately. If the same material is delivered via WWW -service, personalization is much easier.

5. Implications of change

One can use our four-layered model to define an electronic publishing process that either reflects the current practice or defines a desired future process. Whichever the case, the model is likely to change during its lifetime. The following sub-sections discuss what implications a change on one layer will have on the other layers.

5.1. Media object

An alteration of the media object layer may imply two types of changes on the media product layer. First, a new media object format without descriptive metadata results in non-usable content. In such case the media product requires product development effort to cope with the new type of format. Second, a new format containing metadata may be manageable with the production system, but requires conversion to another format before it can be presented on any medium.

A new object format may also provide an opportunity for process optimization. Should the new format be more generic than the existing ones, the adoption of the new format may result in resource savings during the process. For example, a new picture format may allow automated conversion to multiple formats that were previously prepared manually.

5.2. Media product

A media product may be altered in different ways. For example, media objects for the media product may be selected based on some new criteria. This requires that metadata descriptions of the media objects contain information necessary for the selection. A media product may also use the media objects differently. This may either require that media object layer provides an object in suitable formats or that there is an additional conversion process from existing media object formats to the new formats.

5.3. Technology platform

A change in the technology platform enables new capabilities. Support for the new capabilities requires product development work. In addition, the media product layer may require changes to adapt to the new capabilities. For example, given a new image format X, a product development effort is needed to create the product templates to convert images to format X or even to produce new authoring tools to create images in the new format.

5.4. Media

A change of the medium layer implies new opportunities and challenges similar to the technology platform. The set of technologies supported by a new medium can be the same as the technology platform for an already supported medium. In this case the publisher can easily adapt his content to the new medium without changes in the media products. For example, if we have a new laptop computer with support for a set of technologies X and the set Y of technologies used in the production process are supported by X, we can publish for the new laptop without a need to touch the technology platform.

6. Observations

The authors have worked extensively with Finnish media industry while developing and verifying the ideas presented in this paper. The actual work consisted of analysis and development of production processes and content development tools in the field of online and CD-ROM publishing. The following list introduces some of the issues and key findings observed during that work.

- Companies are still following the traditional one-way path from authoring to publishing the media product. This makes it difficult to add interactivity and personalization of media products.
- Metadata, which is the key to media flexibility, is under-utilized. Companies have interest in using metadata, but they lack suitable methods, tools, resources, and process. Even if metadata is created, this work is performed only after the actual content development is completed. This adds unnecessary steps to the overall process.
- The ideas of re-use and media flexibility by separating presentation, structure, and technology platform are not yet fully understood and utilized.

Although these ideas raise interest in the companies, a lot of work is needed before those ideas will be taken into practice.

- Experts perform content authoring with dedicated tools. Actual publication of the content, however, uses a number of complex tools developed and maintained by only one or two key persons. These tools have originally been developed as temporary solutions, but are still in use due to the lack of better alternatives. The tools are vulnerable, lack integration capabilities, and are difficult to maintain, understand, and improve.
- Proper process tools are not used for electronic publishing. Process management and development are still in early stages.
- Certain media products require a lot of unnecessary manual work. Some products can be automatically created while other layout-intensive products require a lot of manual effort to be published. In the latter case it is important to realize, which part of the manual work is truly needed and which could be automated.
- Media products have varying level of platform independence. Some of them have medium-specific additions to content and their own layout, whereas some of them use media objects directly without any platform specific characteristics.

7. Future work

This paper presents an electronic publishing process decomposition model as one entity. We acknowledge the fact that not all companies perform all of the steps. First, our model can be used to *analyze* existing or new publishing processes. Second, our model is useful to *categorize* companies involved in the electronic publishing process based on the steps they implement.

The model does not explicitly support iterations, although they can in most cases be incorporated into the model. Better support for iterative electronic publishing process may be included in the future versions of the model.

The model could acknowledge virtual and physical characteristics of media products. The production and publication of physical products have a significant impact on the cost and flexibility of the process.

The model should also acknowledge better the users of the content. The model should translate customer needs into requirements on different layers of the model. It might also be possible to include customers as the 5^{th} level on top of the media platform layer.

The model requires further verification and practical examples via implementations. This information could improve the model and assist in implementing it in different environments.

Financial aspects could be more emphasized in our work. This could be done for example by incorporating a cost model such as [1] to the framework. A publisher could then evaluate the costs involved with different types of processes before implementing them in practice.

The process aspects of the model could be improved by applying process development principles such as the Capability Maturity Model (CMM) approach [8]. In this way we could add quantitative measurement, feedback, and process improvement to the process model.

8. Conclusions and Recommendations

In this paper we have presented two abstract models that assist in understanding the electronic publishing of multiple products on multiple media.

Our first model divides electronic publishing into four layers. The layers are media, technology platform, media product, and media object. This first model is used to understand the components and their dependencies in electronic publishing. The second model decomposes electronic publishing into process steps. These steps can be categorized under product development, authoring, content composition, and delivery. This model describes the dependencies between the steps and enables us to analyze and optimize a given electronic publishing process.

By combining these two models we are able express the complexity of interaction within the electronic publishing. These two models also make it easier to understand concepts such as re-usability and flexibility with their practical implications.

We have also presented optimization principles to efficiently publish multiple media products on multiple media. Our key recommendations are:

- Try to minimize the required routine manual work in the electronic publishing process. This can be achieved, for example, by using metadata descriptions and content formats that can be automatically converted to other formats.

- Separate authoring, content composition, and delivery of the content. This means content authoring should be completed before any layout or selection related operations.

- Keep electronic publishing and product development activities separate due to their different nature. However, ensure communication between the activities.

- Separate structure from presentation and avoid changing the technology platform while altering other layers. If a media product or a technology platform is changed and their interface is not well defined, changes are probably required on both layers.

- If economically and technologically feasible, isolate the features of media objects and media products in templates. Create these templates before production is started. Templates will enable re-use and automation of the production.

9. Acknowledgements

This research has been financed by Alma Media corporation, Nokia Research Center, Nokia fund, WSOY, Sonera, TeamWare, ICL, Sonera fund, and Helsinki Graduate School in Computer Science and Engineering that is here gratefully acknowledged.

10. Biography

Sami Jokela is a project manager and researcher at the Helsinki University of Technology / IIA Research Centre. His project team researches personalized electronic publishing. Their project, SmartPush, is conducted together with a number of industry partners and has a strong focus on using developed methods in practise. Prior to joining his current project, Jokela worked as a consultant at Andersen Consulting concentrating on both technology and business process issues.

Janne Saarela is a visiting scientist at the W3C. Having previously worked at CERN and at the Helsinki University of Technology on document management issues, he is now preparing his Ph.D on leveraging electronic publishing with metadata.

References

[1] Bot, M., Burgemeester, J.and Roes, H, "The Cost of Publishing an Electronic Journal. A general model and a case study", D-Lib magazine. November, 1998. http://www.dlib.org/dlib/november98/11roes.html (1998).

[2] Bray, T., Paoli, J. and Sperberg-McQueen, C. M, "Extensible Markup Language (XML) 1.0", W3C Recommendation 10-February-1998, http://www.w3.org/TR/1998/REC-xml-19980210 (1998).

[3] Green, D.G, "A General Model for On-line Publishing", Proceedings of the AUUG 96 & Asia Pacific World Wide Web 2nd Joint Conference, http://www.csu.edu.au/special/ auugwww96/proceedings/green/green.html (1998).

[4] Hoschka, P, "Synchronized Multimedia Integration Language (SMIL) 1.0 Specification", W3C Recommendation 15-June-1998, http://www.w3.org/TR/REC-smil/ (1998).

[5] International Organization for Standardization, "Multimedia Hypermedia Experts Group (MHEG) Part 5 Conformance Model. ISO/IEC 13522-5:1997 (1997).

[6] International Organization for Standardization, "Standard Generalized Markup Language (SGML)", ISO 8879:1986 (1986).

[7] Meersman, R., Tari, Z. and Stevens, S, "Database Semantics. Semantic Issues in Multimedia Systems", Kluwer Academic Press, Massachusetts, U.S.A (1999).

[8] Paulk, M., Curtis, B., Chrissis, M. B. and Weber, C. V, "Capability Maturity Model, Version 1.1", IEEE Software, Vol. 10, No. 4, July 1993, pp. 18-27, ftp://ftp.sei.cmu.edu/pub/cmm/Misc/cmm.pdf (1993).

[9] Sheth, A. and Klas, W, "Multimedia Data Management. Using Metadata to Integrate and Apply Digital Media", McGraw-Hill, New York, U.S.A (1998).

[10] Swick, R, "Metadata Activity Statement", W3C Technology and Society domain, http://www.w3.org/Metadata/Activity.html (1998).

[11] Van Den Hamer, P. and Lepoeter, K, "Managing Design Data: the Five Dimensions of CAD Frameworks, Configuration management, and Product Data Management", Proceedings of the IEEE, 84 (1), pp. 42-56 (1996).