Canonical Processes of Semantically Annotated Media Production

Lynda Hardman*
CWI
P.O. Box 94079, 1090 GB
Amsterdam, The Netherlands

Željko Obrenović
CWI
P.O. Box 94079, 1090 GB
Amsterdam, The Netherlands

Frank Nack
University of Amsterdam
Kruislaan 419, 1098 VA
Amsterdam, The Netherlands

Brigitte Kerhervé
Département d'Informatique
UQAM, C.P. 8888
Montreal, QC, Canada

Kurt Piersol
Ricoh Innovations, Inc.
2882 Sand Hill Road / 115
Menlo Park, CA, USA

ABSTRACT

While many multimedia systems allow the association of semantic annotations with media assets, there is no agreed-upon way of sharing these among systems. As an initial step within the multimedia community, we identify a small number of fundamental processes of media production, which we term canonical processes. We specify their inputs and outputs, but deliberately do not specify their inner workings, concentrating rather on the information flow between them. We thus identify a small set of building blocks that can be supported in semantically aware media production tools. The processes are identified in conjunction with a number of different research groups within the community who supply, in companion papers, descriptions of existing systems and a mapping to them. We give a basic formalization of the processes and discuss how this fits with other formalization endeavours. We present a number of frequently asked questions during the development of the model and this special issue.

1. INTRODUCTION

There is substantial support within the multimedia research community for the collection of machine-processable semantics during established media workflow practices [4, 5, 10, 12, 14]. An essential aspect of these approaches is that a media asset gains value by the inclusion of information about how or when it is captured or used, and how it is manipulated and organized. For example, metadata captured from a camera on pan and zoom information can be later used for supporting an editing process [3]. Though the combination of description structures for data, metadata and work processes is promising, current approaches share an essential flaw, namely that the descriptions are not sharable. The problem is that each approach provides an implicit model for exchanging information that serves the particular functionality and process flow addressed by a particular environment.

Working with multimedia assets involves their capture, annotation, editing, authoring and/or transfer to another application. Existing tools tend to concentrate on a single aspect of media production to reduce the complexity of the interface and to simplify implementation. Very often there is no consideration for input requirements of the next tools down the line. While these tools are tailored to support a specific task, they have the potential for adding semantic annotations to the media asset, describing relevant aspects of the asset and why it is being used for a particular purpose. These annotations also need to be included in the information handed on to the next tool

To address these issues of metadata capture, preservation and exchange, we propose an approach to improving interoperability of (semantically-rich) multimedia systems based on a model of canonical processes of media production. Our hypothesis is that the existence of such a unified model can facilitate interoperability among software built by the multimedia community. The primary purpose of the special issue is to gain community agreement on a small number of very basic processes involved with capturing, interpreting and annotating media. Our aim is to establish clear interfaces for the information flow across processes among distinct production phases so that compatibility across systems from different providers can be achieved. We see this as a first step towards a longer term goal — namely, to provide agreed-upon and rigorous descriptions for exchanging semantically annotated media assets among applications.

Our goal is to encourage system creators to provide the outputs we identify when the processes are supported within their system. We hope that in this way the multimedia community will be able to strengthen itself by providing not just single process tools, but enabling these to belong to a

*Lynda Hardman is also affiliated with the Technical University of Eindhoven.

(global) suite of mix and match tool functionality. Another important role of identifying canonical processes is providing a clear definition of concepts so that researchers can coordinate their efforts and instructors can explain concepts to students. As articles in this special issue show, authors often use different terminology to describe similar functionality.

The canonical processes should not be viewed as prepackaged and detailed, ready to be implemented. Our goal is rather to analyse existing systems to identify and generalize functionality they provide and, on the basis of the processes supported within the system, determine which inputs and outputs should be made available.

In this special issue the focus is not on rigorous formalisation of canonical processes, but on their identification and mapping to real systems. In this paper we identify and define a number of canonical processes of media production and give an initial formal description. The other papers in the special issue describe existing systems, the functionality they support and a mapping to the identified canonical processes. These companion system papers are used to validate the model by demonstrating that a large proportion of the functionality provided by the systems can be described in terms of the canonical processes, and that all the proposed canonical processes are supported in more than a single existing system. We do not claim that the canonical processes are able to describe all desired functionality, and discuss issues with extending the canonical processes later in the paper.

This special issue is the result of discussions with many researchers and practitioners in the community. This was initiated in a workgroup on "Multimedia for Human Communication" at a Dagstuhl seminar 05091[1] with a follow-up workshop at ACM Multimedia 2005[2] on "Multimedia for Human Communication - From Capture to Convey". Our goal with these discussions and the open call for papers for the special issue has been to establish community agreement on the model before presenting it here.

In the next section, we briefly describe the companion system papers and the application areas they cover. We then discuss our requirements for describing a single canonical process. The main section of the paper, the definition of the canonical processes, identifies and describes a number of processes we see as being canonical to media production. We then discuss what we have achieved and still need to achieve with the current state of the canonical process descriptions.

2. COMPANION SYSTEM PAPERS

Note to reviewers: This section gives an overview of the papers selected for inclusion in the special issue. Since these are currently under review we have omitted the specific references.

Papers in this special issue come from very diverse areas. They include feature extraction systems, professional news productions systems, new media art, hyper-video production, photo book production, non-linear interactive systems, systems for production of media abstracts, and ambient multimedia systems with complex sensory networks.

3. DESCRIBING CANONICAL PROCESSES

We have chosen the term canonical to indicate that a process cannot be subdivided into more simple constituent processes: *canonical: reduced to the simplest and most significant form possible without loss of generality*[3]. A process is defined in terms of its inputs and outputs and is independent of whether the process can, or should, be carried out by a human or a machine. This allows for a gradual shift of the processing burden from human to machine as technology develops.

To assist in describing the processes we use the Unified Modeling Language (UML[4]). Although UML has limitations, it is a widely adopted standard, familiar to many practitioners, widely taught in undergraduate courses, and supported by many books and training courses. In addition, many tools from different vendors support UML. We hope that our use of UML will facilitate the use of multimedia knowledge by software engineers.

In this section we describe a metamodel that defines a vocabulary of modelling primitives used to describe the media production processes. We then introduce a number of UML extensions based on the metamodel, which are used to describe the canonical processes in the following section.

3.1 Metamodel of Canonical Processes

Figure 1 shows the metamodel where we formally describe basic concepts of media production processes. A *media production process* can be complex or basic. A *complex process* is composed of several basic or complex processes. A *basic process* is represented as a unit that cannot be decomposed into other processes. Each media production process is defined by:

- input that it receives from the real world, such as thoughts of the authors, or user input;
- input that it receives from other processes, such as existing annotations or captured media;
- process and/or real-world artifacts it produces, and
- actors that it involves, such as editor, operator, or designer. An actor can also be some other system or a piece of software.

Output of each process can be the input for other processes, while some processes, such as premeditation, receive input only from the real world. An artifact produced by processes can be atomic, or composite. An atomic artifact defines an artifact that, for a given level of abstraction, cannot be decomposed. We define an *atomic artifact* as:

[1] http://www.dagstuhl.de/en/programm/kalender/semhp/?semnr=05091

[2] http://www.cwi.nl/~media/conferences/mhc05/mhc05.html

[3] http://wordnet.princeton.edu/perl/webwn?s=canonical

[4] http://www.uml.org/

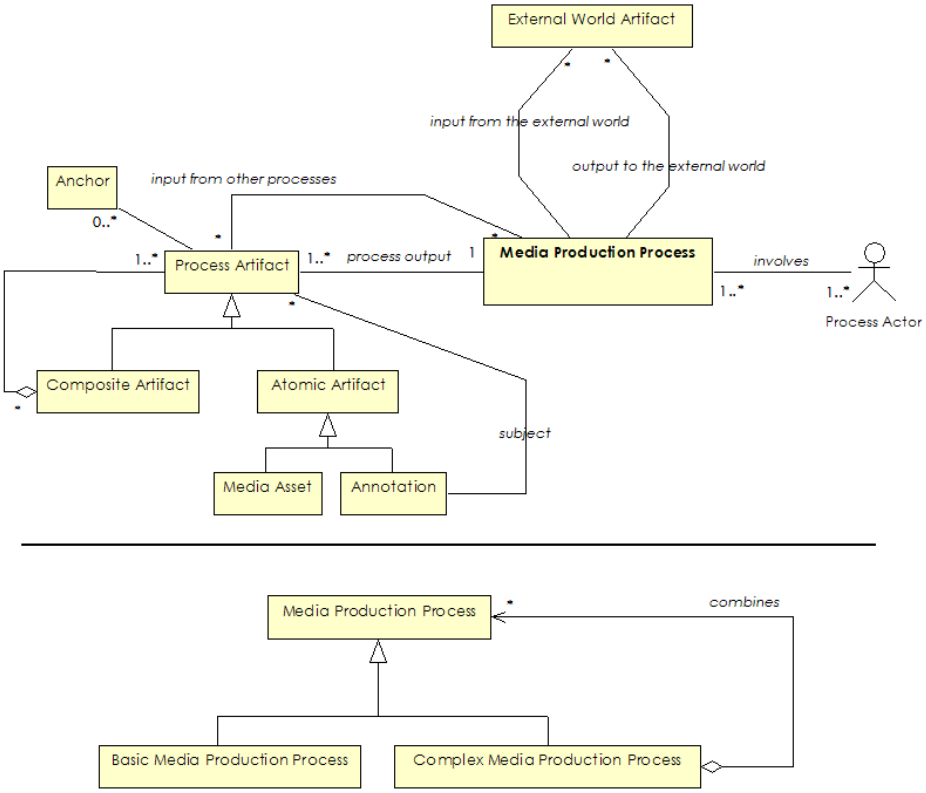

Figure 1: The metamodel of canonical processes of media production.

- Media asset[5], such as captured video or an image. Usually it contains "raw" data recorded with some sensor technology, such as camera or microphone, but it can also be a product of editing or drawing programs, or the result of transformation of existing media assets.

- Annotation[6], any pertinent information, whether denotative or connotative, such as a description of what is represented in a media asset, or how the media asset is created. The subject of the annotation can be any process artifact (or part of the artifact identified by an anchor [7, 8]), including other annotations. Optionally, an annotation could be based on terms defined by some schema, e.g. in MPEG-7 [13] or OWL [16]. We make no restrictions or assumptions on the semantics and format of an annotation.

A *composite artifact* is composed of other atomic or composite artifacts. We specify no further semantics of the composition[7].

[5]This is equivalent to the Dexter [7] and Amsterdam Hypermedia Model (AHM) content of the atomic component, `http://www.cwi.nl/~lynda/thesis/A1.pdf`

[6]This is equivalent to the Dexter [7] and AHM attributes of the atomic component, `http://www.cwi.nl/~lynda/thesis/A1.pdf`

[7]The AHM specifies two structural compositions, temporal and atemporal, which describe presentation oriented composition, `http://www.cwi.nl/~lynda/thesis/A1.pdf`

Note that terms media asset and annotation represent roles that a particular information object plays for a process. For different processes the same information object may have different roles.

3.2 UML Extensions for Description of Canonical Processes

We specify a UML profile where we introduce several UML extensions based on the proposed metamodel. With these extensions, we can describe processes of media production at different levels of abstraction, with various levels of detail. UML includes a formal extension mechanism to allow practitioners to extend the semantics of the UML. The mechanism allows us to define stereotypes, tagged values and constraints that can be applied to model elements[8]. In this paper, we describe only the stereotypes we use. Table 1 shows some of introduced UML class and association stereotypes. We use these stereotypes to describe the canonical processes in the next section.

4. CANONICAL PROCESSES OF MEDIA PRODUCTION

[8]A stereotype is an adornment that allows us to define a new semantic meaning for a modeling element. Tagged values are key value pairs that can be associated with a modeling element that allow us to "tag" any value onto a modeling element. Constraints are rules that define the well-formedness of a model. They can be expressed as free-form text or with the more formal Object Constraint Language *OCL*.

Name	Type	Description
<<process>>	class stereotype	Describes a process.
<<process artifact>>	class stereotype	Describes any process artifact.
<<media asset>>	class stereotype	Describes a media asset as a basic artifact.
<<annotation>>	class stereotype	Describes an annotation as a basic artifact.
<<composite artifact>>	class stereotype	Describes a composite artifact.
<<process actor>>	class stereotype	Describes a process actor.
<<external world artifact>>	class stereotype	Describe entities from external world.
<<input>>	association stereotype	Connects processes with input artifacts.
<<output>>	association stereotype	Connects process with output artifacts.

Table 1: UML stereotypes used to describe processes of media production.

Based on examination of existing multimedia systems, we have identified nine canonical processes of media production. Every process introduced into our model has at least several instances in existing systems. Our model, therefore, does not contain processes that are specific for particular systems. In the following sections we describe in detail these nine processes:

- *premeditate*, where initial ideas about media production are established,
- *create media asset*, where media assets are captured, generated or transformed,
- *annotate*, where annotation is created,
- *package*, where process artifacts are logically and physically packed,
- *query*, where a user retrieves a set of process artifacts based on a given query,
- *construct message*, where an author specifies the message they wish to convey
- *organize*, where process artifacts are organized according to the message,
- *publish*, where final content and user interface is created, and
- *distribute*, where final interaction between end-users and produced media occurs.

For each process, we give a detailed explanation and state its inputs, outputs and involved actors. A preliminary, less formal, diagram can be found at the Dagstuhl web site[9]. While we use a single name to name each of the processes, these are meant to be used in a very broad sense. The textual description of each one specifies the breadth of the process we wish to express.

In this paper we describe only basic processes. Particular real systems, described in the companion system papers, describe composite processes that combine several canonical processes.

Note to reviewers: The final version of this paper will include references to the companion system papers throughout this section.

[9]`http://www.dagstuhl.de/files/Proceedings/05/05091/05091.PiersolKurt1.Slides.pdf`

4.1 Premeditate

Any media creation occurs because someone has made a decision to embark on the process of creating — whether it be image capture with a personal photo camera, drawing in a drawing tool, professional news video, an expensive Hollywood film or a security video in a public transport system. In all cases there has been premeditation and a decision as to when, how and for how long creation should take place.

In all these cases what is recorded is not value-free. A decision has been made to take a picture of this subject, conduct an interview with this person, make this take of the chase scene or position the security camera in this corner. Already there are many semantics that are implicitly present. Who is the "owner" of the media to be created? Why is the media being created? Why has this location/background been chosen? Whatever this information is, it should be possible to collect it and preserve it and be able to attach it to the media that is to be created. For this we need to preserve the appropriate information that can, at some later stage, be associated with one or more corresponding media assets.

Figure 2 shows a UML class diagram of the premeditate process described in terms of the metamodel. The input to this process are ideas, decisions, and artifacts from outside the system. The output is a set of premeditate artifacts, which are more typically annotations.

4.2 Create Media Asset

After a process of premeditation, however short or long, at some point there is a moment of media asset creation. Some device, for example, is used to collect images or sound for a period of time, be it photo or video camera, scanner, sound recorder, heart-rate monitor, MRI etc.

Note that in this process, we do not restrict creation of a media asset to only newly recorded information. Media assets can also be created in other ways. For example, images can be created with image editing programs or generated by transforming one or more existing images. The essence is that a media asset comes into existence, we are not interested in the method of creation *per se*. If the method is considered as significant, however, then this information should be recorded as part of the annotation.

Figure 3 shows a UML class diagram of the create media asset process described in terms of the metamodel. The input to the capture process is a collection of annotations, for example, information available from the premeditation process, and/or the message construction process. As a result of the create media asset process we have new media assets. This process usually involves one or more creation actors,

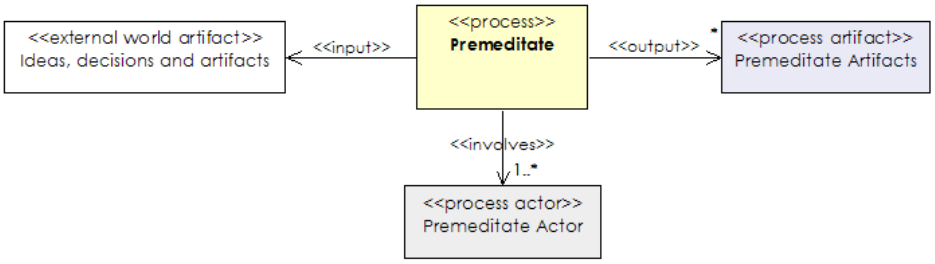

Figure 2: A class diagram describing the premeditation process.

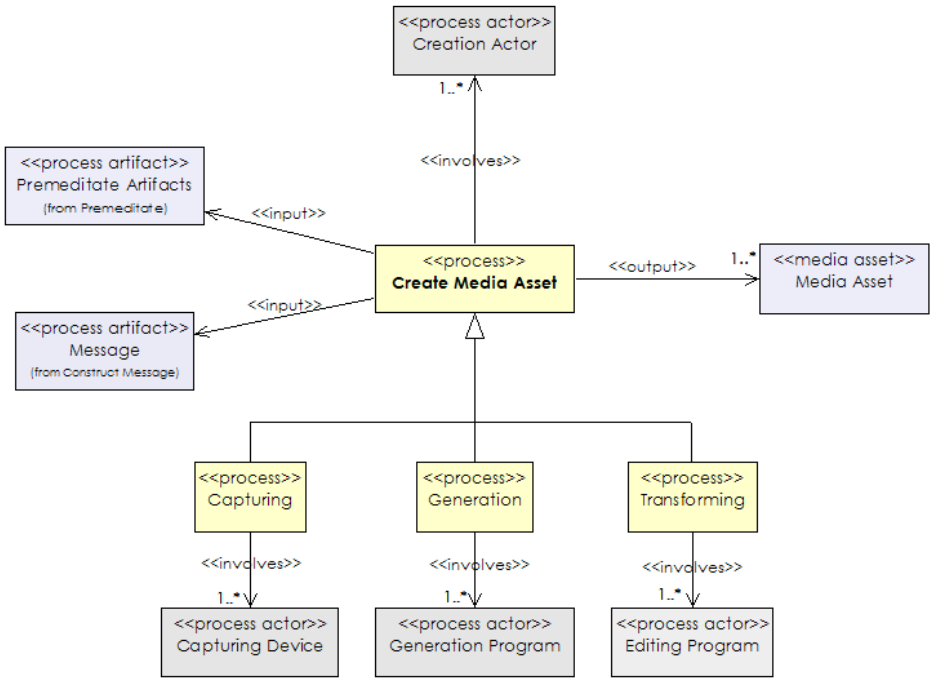

Figure 3: A class diagram describing the media asset process.

such as a camera person.

4.3 Annotate

The annotate process allows extra information to be associated with any existing process artifact. The term annotation is often used to denote a single human user adding metadata to facilitate search. Here we view annotation as the broader process of adding more easily machine-processable descriptions of the artifact.

The annotations need not be explicitly assigned by a user, but may be assigned by an underlying system, for example by supplying a media asset as input to a feature analysis algorithm and using the extracted result to annotate the media asset. We make no distinction whether annotations are selected from an existing vocabulary or machine generated. If deemed relevant, the identity of the human assigner or the algorithm can be recorded in the annotation [1].

We do not prescribe the form of annotations, but require that they can be created and associated with one or more artifacts. We also do not impose limitations on the structure of annotations, due to the high diversity of annotation formats in practice. In most semantically-rich systems, however, the structure of an annotation may include a reference to a vocabulary being used, one of the terms from the vocabulary plus a value describing the media asset (we call this process a semantic annotation).

The annotation can refer to any artifact as a whole, but the annotation could also be more specific. In this case, an anchor mechanism is needed to refer to the part of the media asset to which the annotation applies [7]. An anchor consists of a media independent means of referring to a part of the media asset and a media-dependent anchor value that specifies a part of the media asset. For example, for an image this could be an area, for an object in a film a time-dependent description of an area of the image. For further discussion on anchor specifications see [8] p53.

Figure 4 shows a UML class diagram of the annotate process described in terms of the metamodel. The input to this process is a process artifact plus an annotation supplied by a human or system. The output is a process artifact including the extra annotation. In a semantic annotate process

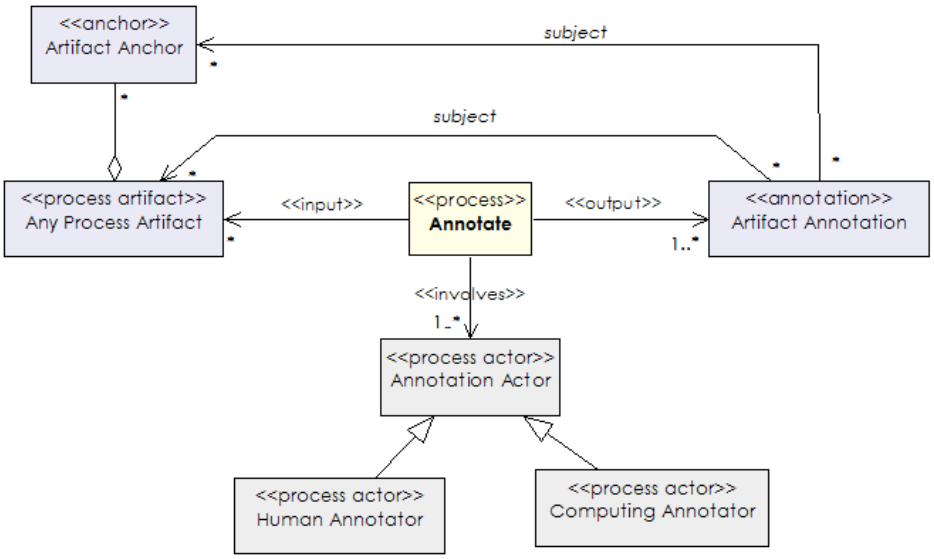

Figure 4: A class diagram describing the annotation process.

specialisation, the annotation itself may contain a reference to a vocabulary, element and/or value used in the annotation (Figure 5). The process involves one or more human or computing annotators.

4.4 Package

The process of packaging provides a message independent grouping of artifacts, and assigning this grouping an identity so that it can be retrieved as a unit. Grouping produces a unit, called a multimedia component, from multiple process artifacts. Components can be organized in a hierarchical fashion, i.e. one component can encapsulate other components. We make a distinction between physical and logical packaging. In some cases physical and logical packaging are inseparable, for example, when we put an image in one file, while organization of file directories can reflect the structure of presentation. However, in many situations physical and logical packaging are not equivalent. For example, SMIL presentation logically presents one unit, but links media components physically packaged in many files in a distributed environment. On the other hand, a multimedia database can physically be packaged in one file, but contain many logical units. The component identity is equivalent to the component identity specified in hypertext literature [7, 8].

Figure 6 shows a UML class diagram of the package process described in terms of the metamodel. The input to this phase is a set of process artifacts and optionally input given by user, such as an identifier of created package. The output is a multimedia package that physically or logically groups input artifacts into one unit, and optionally defined anchors.

4.5 Query

The query process selects a number of artifacts from a repository of artifacts. Up until now the processes we describe concentrate on creating, storing and describing primarily media assets. These are needed for populating a media repository. Note that our definition of media repository does not necessarily imply the existence of a complex storage infrastructure, but we assume that systems have a repository where they keep media assets and other artifacts, in the most simple case a hierarchically organized file directory structure. Once there is a repository of artifacts (but not before) it can be queried for components whose associated media assets correspond to desired properties.

We do not wish to use a narrow definition of the term "query", but intend to include any interface that allows the artifacts to be searched, using query languages of choice or (generated) browsing interfaces that allow exploration of the content of the archive. It is worth noting that many systems that provide advanced query interfaces, also provide support for other processes. For example, browser interfaces can, in addition to a simple query interface, also organize intermediate results to present them to a user for feedback, and create temporary presentations that are then published and distributed to the user.

A query of the system may be in terms of media assets, or in terms of the annotations stored with the media assets. A query needs to specify (indirectly) the annotation(s) being used, and includes techniques such as query by example. The mechanisms themselves are not important for the identification of the process.

Figure 7 shows a UML class diagram of the query process described in terms of the metamodel. The input to the query process is a set of process artifacts plus a specification of a subset of these. The output is a (possibly empty) set of identified media components corresponding to the specification. The output is often not a set of media assets, but a structural asset that includes references to the media components that contain links to process artifacts.

4.6 Construct Message

A presentation of media assets, such as a film or an anatomy book, is created because a human author wishes to communicate something to a viewer or reader. Constructing the

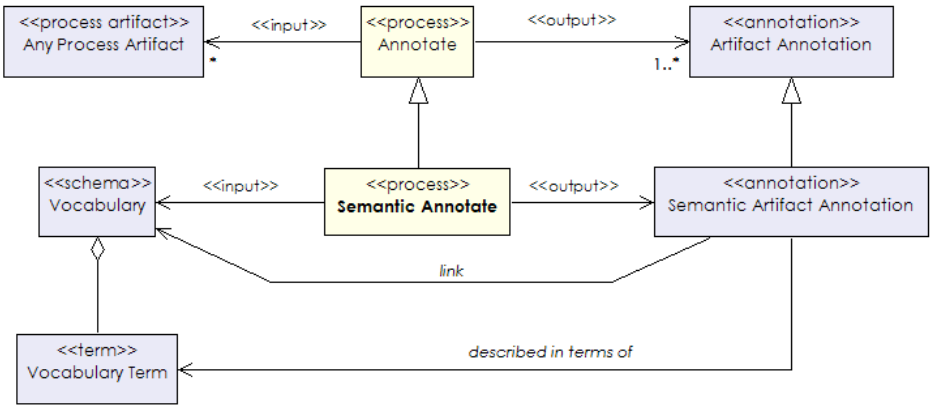

Figure 5: A class diagram describing the semantic annotation process.

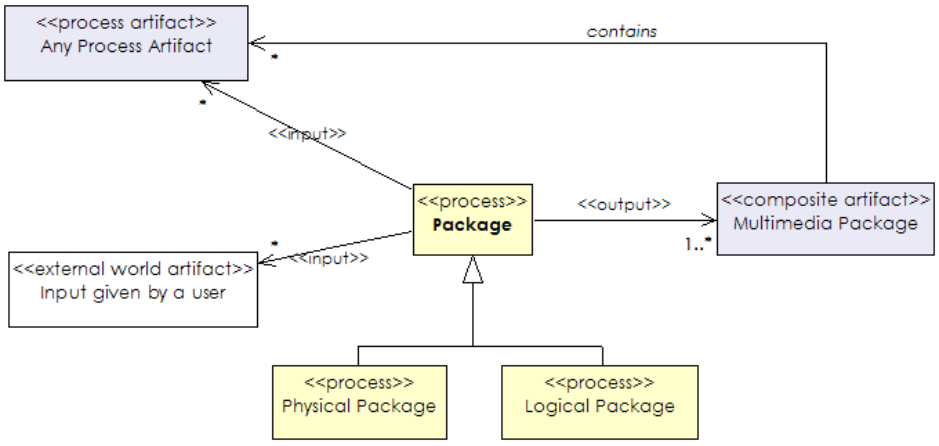

Figure 6: A class diagram describing the package process.

message which lies behind the presentation is most often carried out by one or more human authors. When a viewer watches a film or a reader reads a book then some part of the intended original message of the author will hopefully be communicated. In order to give different processes access to the underlying intent, we include an explicit process which brings a processable form of the message into the system. Just as capturing a media asset is input into the system, so is the specification of the message an author wishes to convey.

In some sense, there is no input into the construct message process. However, the real input is the collection of knowledge and experience in the author her/himself. The output of the process is a description of the intended message. For example, a multimedia sketch system, such as described in [2], allows an author to gradually build up a description of the message. For the message to be machine processable the underlying semantics need to be expressed explicitly.

A query implicitly specifies a message, albeit a simple one, that an author may want to convey, since otherwise the author would not have been interested in finding those media assets. The query is, however, not itself the message that the author wishes to convey.

While we do not exclude this process as being carried out by a system, we expect that, at least in the near future, it will predominantly be carried out by a human user.

In general, we give no recommendation in this paper for the expression of the semantics of the message. We expect that it contains information regarding the domain and how this is to be communicated to the user, but we do not assign anything more than a means of identifying a particular message.

Figure 8 shows a UML class diagram of the construct message process described in terms of the metamodel. The input to this process are ideas and decisions from the external world, and the output is a message.

4.7 Organise

While querying allows the selection of a subset of media assets, it imposes no explicit structure on the results of one or more queries. The process of organisation is to create some document structure for grouping and ordering the selected media assets for presentation to a user. How this process occurs is, again, not relevant, but includes, for example, the linear relevance orderings provided by most information retrieval systems. It also includes the complex human process of producing a linear collection of slides for a talk; creating

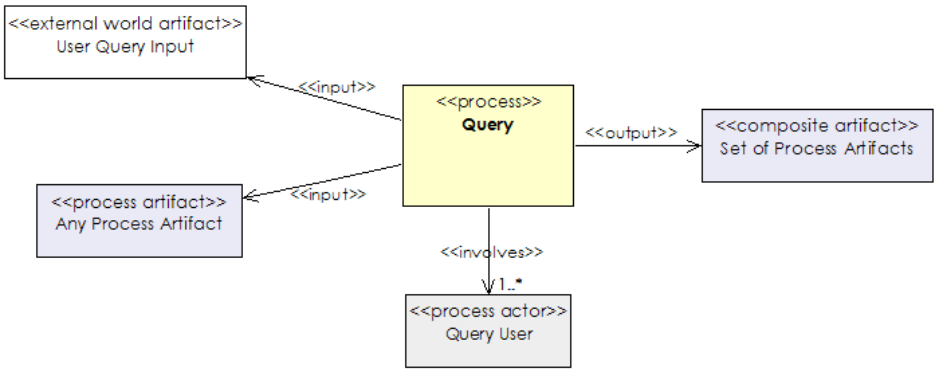

Figure 7: A class diagram describing the query process.

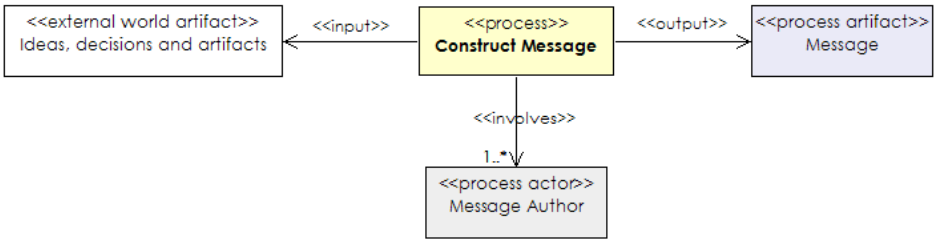

Figure 8: A class diagram describing the construct message process.

multimedia documents for the web; ordering shots in a film; or even producing a static 2-dimensional poster.

The process of organisation is guided by the message (the output of the construct message process). The organization depends on the message and how the annotations of the process artifacts relate to the message. For example, annotations concerning dates could be used to order assets temporally. The resulting document structure may reflect the underlying domain semantics, for example a medical or cultural heritage application, but is not required to. The structure may be colour-based or rhythm based, if the main purpose of the message is, for example, aesthetic rather than informative.

In the arena of text documents, the document structure resulting from organisation is predominantly a hierarchical structure of headings and subheadings. The document structure of a film is a hierarchical collection of shots. For more interactive applications, the document structure includes links from one "scene" to another. In a SMIL [15] document, for example, `par` and `seq` elements form the hierarchical backbone of the document structure we refer to here.

Figure 9 shows a UML class diagram of the organize process described in terms of the metamodel. The input to the organize process is the message plus one or more process artifacts. The output is the document structure, which includes the media assets associated with the substructures.

4.8 Publish

The output of the organize process is a prototypical presentation that can be communicated to an end-user. This serves as input to the publication process which selects appropriate parts of the document structure to present to the end-user. The publication process takes a generic document structure and makes refinements before sending the actual bits to the user. These may include selecting preferred modalities for the user and displayable by the user's device. The resulting presentation can be linear (non-interactive, e.g. a movie) or non-linear (interactive, e.g. web presentation).

Publication can be seen as taking the document structure from the internal set of processes and converting it (with potential loss of information) for external use. Annotations may be added to describe the published document. For example, the device or bandwidth for which the publication is destined. Annotations and alternative media assets may be removed to protect internal information or just reduce the size of the data destined for the user.

Figure 10 shows a UML class diagram of the publish process described in terms of the metamodel. The input to the publication process is a set of process artifacts and a message that guides the organization, and the output is a published document that organizes input process artifacts according to the message.

4.9 Distribute

Created content has to be, synchronously or asynchronously, transmitted to the end-user. This final process involves some form of user interaction and requires interaction devices, while transmission of multimedia data to the user device goes through some of the transmission channels including the internet (streamed or file-based) non-networked medium (such as a CD-ROM or DVD) or even analog recording media (for example, film).

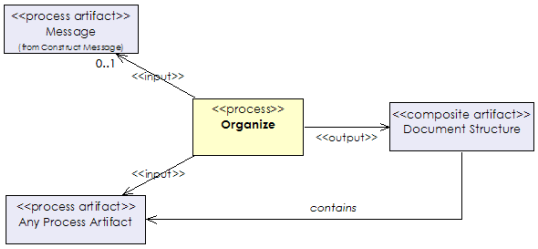

Figure 9: A class diagram describing the organize process.

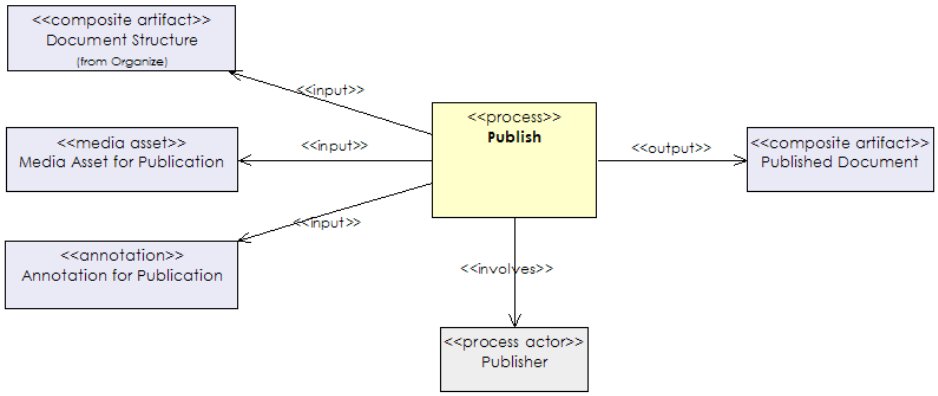

Figure 10: A class diagram describing the publish process.

It is important to note that the term "distribution" in our model has a much broader meaning than in classical linear production of media. It can also be used to describe interactive non-linear productions, such as games or other interactive presentations. The resulting system would implement a complex process including query, organize and publish processes in addition to distribution (end-user interaction). For example, some systems can have a final presentation where the story line depends on user feedback. In this case, the presentation system would include the canonical processes query (to select next part of the story), organize (to organize selected items coherently), publish (to create internal document ready for presentation) and distribute (to present and expose the control interface to user). Thus, for specialised client systems this step is useful. For systems, such as server driven web applications, this step is out of scope.

Figure 11 shows a UML class diagram of the distribute process described in terms of the metamodel. The input to the process is a process artifact representing a published document. This process involves appropriate software/hardware for displaying/playing or interacting with the media assets. The output is the real-time display or projection of the media assets to the end-user, or the creation of a physical carrier, such as DVD or film.

5. DISCUSSION

The canonical processes represent a distillation of our discussions on formulating them to help the multimedia community communicate about their systems at both a descriptive and computational level. During the course of our discussions we encountered questions that arose from multiple authors. We discuss these with the aim of clarifying potential misunderstandings.

We also discuss how our descriptions of canonical processes fit in with more formal descriptions, in particular foundational ontologies, and what the boundaries of our achievements are – in particular we deliberately do not describe the form of the annotations. We see this rather as parallel, but closely related, work.

The section gives an overview of advantages of specifying the canonical processes explicitly, some frequently asked questions and then how the process descriptions can be linked to more formal representations.

5.1 Benefits of descriptions in terms of the canonical processes

Most of the authors in this special issue discussed the potential benefits of canonical processes. Here we summarise the main observations from the authors of the companion papers.

Note to reviewers: These will be supplemented with references to the individual papers for the final version.

5.1.1 Identifying omitted functionality

Identifying and aligning the processes implemented by a particular system with the canonical processes enables a comparison of system functionality. This allows implementors to identify functionality not currently implemented and make informed decisions as to whether to implement the missing functionality in their own system or search for an existing, compatible component.

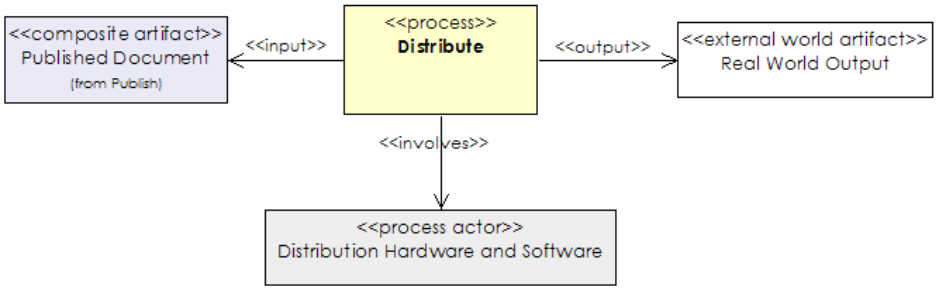

Figure 11: A class diagram describing a distribution process.

The authors of Paper 1 noted that by aligning their system processes with the canonical processes, they were able to better understand the process cycles of their system in the context of the more generalised process cycles of existing systems.

5.1.2 Improving interoperability

Comparing an existing system with the canonical processes also improves interoperability. Defining both application and framework functionalities in terms of canonical processes allows the decoupling of the actual abstraction steps and visualisation and avoids hard-wiring design decisions and context information.

The canonical processes also help to envisage future scenarios where some of the processes within a system could be exchanged with those from other. For example, the publish and distribute processes from other media production systems could be plugged into the Paper 1 system, taking the output of their organise process to generate presentations suitable for different modalities and interaction platforms. The Paper 6 system, for example, could be one of these systems.

5.1.3 Using annotations in different processes

Much work in the multimedia community is geared to analysing visual or audio media and extracting higher-level representations of the features found. These are often geared to facilitating the query process. By identifying a number of other processes, we are now able to discuss the role of annotations in each of these. For example, the organise process can use semantics from both the message and the results retrieved by a query in order to impose more user-friendly ways of grouping and ordering material. The publish process can also benefit from information about the media characteristics (such as bandwidth required, or aspect ratio) for selecting items for a specific platform.

5.2 Frequently asked questions

While preparing the special issue, we had several discussion with authors about how to describe particular elements of their systems in terms of the canonical processes. Here we present the most frequently asked questions.

5.2.1 Complex processes

We encountered requests from authors of companion papers to include more complex processes in the list of canonical processes. The resolution was to better explain that the canonical processes constitute a model, useful for discussing within the community, rather than a prescription for a system architecture. Any particular application may implement functionality that includes multiple canonical processes. Our goal is to make developers aware that this is what they are doing, and if there are intermediate results within the system that these be made available to other modules or systems.

5.2.2 Interaction

The relationship between interaction and the canonical processes is not clear in the first instance. While the canonical processes describe a query process and a distribute process, there is no process termed "interaction" or "user feedback". Our view is that the process of interaction takes place within the context of an application, which may implement one or more of the canonical processes, and that interaction is, and should remain, a complex process including multiple canonical processes. The new media art application discussed in [11] is an example that illustrates this. Users moving through an interactive art installation provide input that is used as a precursor to a query process. This in turn leads to processing within the system, which then takes the results from the query process and, via an organise process, distributes the result back to the user. The interaction cycle takes place on sub-second timescales, during which different canonical processes are invoked.

The process of interaction itself can also be used to contribute to an annotation process. For example, if an application records that many people select the same media asset many times, then this could be captured as an annotation and used to reorder the results on a subsequent organisation of the same query.

5.2.3 Complex artifacts and annotations can be annotated

The composite artifact representing a complex presentation resulting from the organize process can be fed back into any process that accepts a composite artifact as input. This can be used to associate annotations with the complex artifact, such as its author or a version number. For example, a film script is the result of a long premeditation process with a complex structure whose semantics can only be captured with difficulty. The complete script represented by an atomic artifact, however, is a process artifact with which annotations can be associated. An annotation is itself an atomic artifact, hence a process artifact with which anno-

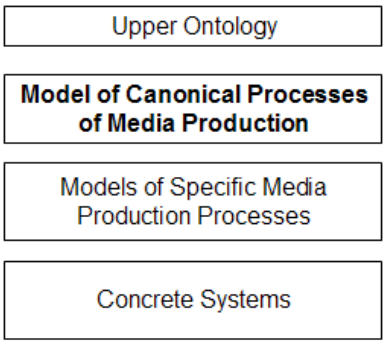

Figure 12: Relation of the model of canonical processes of media production with upper ontologies and specific models and systems.

tations can be associated. For example, the author who assigned the annotation.

5.2.4 *Further specification of canonical processes*

While the canonical processes provide a small set of identified processes common across systems, some authors felt the need to specify more details of a number of the processes for particular situations. While we see the need for agreement on a further level of detail in some of the processes, our goal for this special issue is to obtain community wide agreement on the identification of a small number of processes. A further stage of evolution could be to encourage different communities to agree upon more detailed specifications in specialised areas. We see this as similar to the community process of describing agreed-upon XML schemas or domain-specific ontologies.

5.3 Towards a more rigorous formalization of the model

In this paper we provide a UML description of the processes. While a text-only description is insufficiently precise, we do not want to exclude a large audience by using unfamiliar formal languages. We do, however, see the need for future work to link our current descriptions into a higher-level ontology, and to specify the structure of annotations more precisely.

5.3.1 *Relationship to foundational ontologies*

A foundational ontology, or upper-level, ontology, describes a domain independent vocabulary that explicitly includes formal definitions of foundational categories, such as processes or physical objects. It provides a hierarchy of entities and associated rules (both theorems and regulations) that describe general entities that apply across domains. Describing the processes using a foundational ontology, such as DOLCE [6], provides a solid modelling basis and enables interoperability with other models. DOLCE provides description templates (patterns) for the specification of particular situations and information objects. Figure 12 illustrates the relation of the model of canonical processes of media production and specific models and systems, and with upper ontologies.

A next step in the specification of the ontologies would be to express these as specialisations of the DOLCE model, in particular, the DOLCE situations.

5.3.2 *Semantics of Annotations*

Although out of the scope of this special issue, our effort to identify canonical processes is part of a broader effort to create a rigorous formal description of a high quality multimedia annotation ontology compatible with existing (semantic) web technologies. In particular, we are involved with work on specifying the structure of complex semantic annotations of non-textual data. This has resulted in COMM – A Core Ontology for Multimedia[10] [1] based on the MPEG-7 standard [9] and expressed in terms of OWL [16].

6. CONCLUSION

In this paper we identify a number of canonical processes that are fundamental to different applications of multimedia. This model forms an initial step towards the definition of data structures, which could be accessed and produced by systems created within the multimedia community. In the rest of this special issue we present a number of companion papers that describe existing systems in terms of the canonical processes.

While the canonical processes have been developed within the multimedia community, the semantic web community is also investigating ways to resolve the problem of keeping data and its associated annotations together in everyday desktop environments – a so-called semantic clipboard. Initial work on this has been carried out [**?**], and we see the canonical processes as being helpful in providing a framework for researchers to investigate particular solutions and adapt solutions from neighbouring fields.

The companion papers illustrate the breadth of applicability of the processes, but are not intended to restrict its scope. Many other application areas of semantically annotated media would potentially benefit from their consideration during the design process. For example, universal design[11], advocates *the design of products, services and environments to be usable by as many people as possible regardless of age, ability or circumstance*, is an area that could benefit to a great extent from more explicit management of multimedia and associated annotations **??**. For example, annotations from a premeditate process, such as a script, can provide a basis for creating audio descriptions of movies for blind users, or textual descriptions of audio effects for deaf users. Information about organization of the content can make easier navigation for blind users.

While we have identified a set of canonical processes for semantic multimedia, these have not been developed in isolation. As described in the discussion section, COMM provides a model for describing the annotations mentioned, but not specified, in the canonical processes. Other relevant work on semantically annotating images on the web has been carried out by the W3C Incubator Group on Multimedia Semantics[12]. We see these different threads stimulating and complementing each other towards open web-based

[10] `http://multimedia.semanticweb.org/COMM/`
[11] `http://en.wikipedia.org/wiki/Universal_design`
[12] `http://www.w3.org/2005/Incubator/mmsem/`

data structures and software components for describing and sharing semantically annotated media assets among different platforms.

Acknowledgments

This paper was inspired by Dagstuhl meeting 05091 "Multimedia Research - where do we need to go tomorrow" organized by Susanne Boll, Ramesh Jain, Tat-Seng Chua and Navenka Dimitrova. Members of the "Multimedia for Human Communication" working group were: Lynda Hardman, Brigitte Kerherve, Stephen Kimani, Frank Nack, Kurt Piersol, Nicu Sebe, and Freddy Snijder. We would also like to thank all the authors of the companion system papers, in particular Susanne Boll and Ansgar Scherp, for their feedback on the model. In addition we would like to thank Raphaël Troncy and Jacco van Ossenbruggen for their valuable feedback on this paper.

Parts of this research were funded by the Dutch national BSIK MultimediaN e-Culture, ToKeN2000 CHIME and European ITEA Passepartout and IST K-Space projects.

7. ADDITIONAL AUTHORS

8. REFERENCES

[1] R. Arndt, R. Troncy, S. Staab, and L. Hardman. Adding Formal Semantics to MPEG7: Designing a Well-Founded Multimedia Ontology for the Web. Technical Report KU-N0407, KU and CWI, January 2007.

[2] B. P. Bailey, J. A. Konstan, and J. V. Carlis. Supporting Multimedia Designers: Towards More Effective Design Tools. In *Proc. Multimedia Modeling: Modeling Mutlimedia Information and Systems (MMM2001)*, pages 267–286. Centrum voor Wiskunde en Informatica (CWI), 2001.

[3] S. Bocconi, F. Nack, and L. Hardman. Using Rhetorical Annotations for Generating Video Documentaries. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME) 2005*, July 2005.

[4] M. Davis. Active Capture: Integrating Human-Computer Interaction and Computer Vision/Audition to Automate Media Capture. In *ICME '03, Proceedings*, pages 185–188, July 2003.

[5] C. Dorai and S. Venkatesh. Bridging the Semantic Gap in Content Management Systems - Computational Media Aesthetics. In C. Dorai and S. Venkatesh, editors, *Media computing - Computational media aesthetics*, pages 1–9. Kluwer Academic Publishers, June 2002.

[6] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening Ontologies with DOLCE. In *EKAW '02: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, pages 166–181, Siguenza, Spain, October 01–04, 2002. Springer-Verlag.

[7] F. Halasz and M. Schwartz. The Dexter Hypertext Reference Model. *Communications of the ACM*, 37(2):30–39, February 1994. Edited by K. Grønbæck and R. Trigg.

[8] L. Hardman. *Modelling and Authoring Hypermedia Documents*. PhD thesis, University of Amsterdam, 1998. ISBN: 90-74795-93-5, also available at http://www.cwi.nl/∼lynda/thesis/.

[9] ISO/IEC. Overview of the MPEG-7 Standard (version 6.0). ISO/IEC JTC1/SC29/WG11/N4980, Pattaya, December 2001.

[10] R. Jain. Experiential Computing. *Communications of the ACM*, 46(7):48–55, July 2003.

[11] B. Kerhervé, A. Ouali, and P. Landon. Design and Production of New Media Artworks. In *Proceedings of the ACM Workshop on Multimedia for Human Communication - From Capture to Convey (MHC 05)*, November 2005.

[12] H. Kosch, L. Böszörményi, M. Döller, M. Libsie, P. Schojer, and A. Kofler. The Life Cycle of Multimedia Metadata. *IEEE MultiMedia*, (January-March):80–86, 2005.

[13] F. Nack and A. T. Lindsay. Everything You Wanted to Know About MPEG-7: Part 1. *IEEE MultiMedia*, pages 65–77, July - September 1999.

[14] F. Nack and W. Putz. Designing Annotation Before It's Needed. In *Proceedings of the 9th ACM International Conference on Multimedia*, pages 251–260, Ottawa, Ontario, Canada, September 30 - October 5, 2001.

[15] W3C. Synchronized Multimedia Integration Language (SMIL 2.0) Specification. W3C Recommendation, August 7, 2001. Edited by Aaron Cohen.

[16] W3C. Web Ontology Language (OWL) - Overview. W3C Recommendation, 10 February 2004.