# Advene: an open-source framework for integrating and visualising audiovisual metadata

Olivier Aubert
LIRIS UMR 5205 CNRS
Lyon 1 University
F-69622 Villeurbanne, France
olivier.aubert@liris.cnrs.fr

Yannick Prié
LIRIS UMR 5205 CNRS
Lyon 1 University
F-69622 Villeurbanne, France
yannick.prie@liris.cnrs.fr

## ABSTRACT

The open-source Advene prototype offers a framework for integrating and visualising audiovisual metadata. It allows users to define by themselves, according to their specific tasks, the structure of the metadata as well as the different ways in which it should be visualised. By storing metadata and visualisation specifications independently from the audiovisual document, it allows to share analyses and comments on any audiovisual document. Its open nature and simple principles make it an ideal testbed for experimentation with new audiovisual metadata interaction modalities.

## Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Video; H.5.4 [**Hypertext/Hypermedia**]: Architectures, Navigation

## Keywords

Annotation, Advene, Hypervideo, Document Template, Audiovisual Information Visualisation, Sharing, Time and synchronisation

## General Terms

Design,Experimentation

## 1. PRINCIPLE

The Advene project aims at developing an open-source framework for hypervideo engineering, that allows to 1/ annotate audiovisual documents, i.e. to associate information to specific fragments of a video; 2/ provide augmented visualisations of the video that use the annotation structure; 3/ exchange the annotations and their associated visualisation modes independently from the original video, as documentary units called *packages*.

The goal of the project is to foster innovative uses of audiovisual material, allowing users to quickly experiment with new ideas, based on existing or specifically created metadata. For this, we acknowledge the tremendous importance of metadata, on which we focus our reflection. The Advene project tries to integrate three steps of the lifecycle of audiovisual metadata: its creation and evolution, its visualisation, and its exchange.
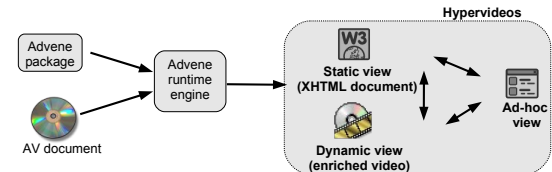
**Figure 1: The Advene principle**

The underlying principle of Advene, as illustrated in figure 1, is to bundle both metadata and their visualisation specification in a single document, called a package. The package can be shared by various means (e-mail, downloaded from a web server...), independently from the audiovisual document itself, and reused by other persons. A basic use of a package simply consists in using its metadata to navigate in the audiovisual document, or to generate new hypermedia documents through views defined by the package author. However, since the user gets a copy of metadata and view definitions, a more advanced and active usage may involve modifying or augmenting either the metadata or its visualisation specifications. The user can for instance create new metadata and visualise it with existing views, or create new view definitions in order to produce a different rendering of the same metadata.

Let us consider for example a community of moviegoers engaged in a discussion about Murnau's Nosferatu movie, wishing to thoroughly analyse how horror is conveyed in it. One of the persons uploads on a webserver a package containing a shot segmentation of the movie, and a selection of interesting scenes. In order to ease their visualisation, the user provides a generated hypertext view that comments the interesting scenes with direct links to the corresponding movie sequences, thus offering a kind of table of contents. Another user extends this work by indexing the textual panels that appear in the movie, accompanied with their translation in french and german. He also provides a hypertext view that displays the text panels linked to the video, allowing to do a full-text search in them. Moreover, he defines an augmented visualisation of the video that captions the movie with the translated french panels. All this metadata and accompanying visualisations is made available to the community and serves as a reference for their future discussions.

To provide these features, the Advene project defines a data model, based on the hypervideo model described in [1]. The model features three main elements: the annotation

structure (structured annotations and relations), the views (specifying how to render the annotation structure together with the audiovisual document) and the queries (allowing to dynamically select elements from the model). Annotations are pieces of information of any type (simple text for basic needs and active reading, audio comments, pdf documents, etc.) that are linked to a specific spatio-temporal fragment of the audiovisual document. The current implementation provides three types of views: adhoc views, which are customizable GUI components, static views, that use templates to render XHTML documents, and dynamic views, that allow to enrich the play of a movie with supplemental information, interaction possibilities, etc.

## 2.  THE GLOBAL ARCHITECTURE

The Advene prototype is a cross-platform, open-source software (licensed under the terms of the GNU GPL), reusing off-the-shelf components (video player, web server, templating system...). The prototype and accompanying examples are freely downloadable from http://liris.cnrs.fr/advene/. Illustrations for this article are taken from the Nosferatu example package, that the reader is invited to download from the website in order to explore the features of Advene.

As presented in figure 2, the Advene application embeds and controls a video player, as well as a small webserver that serves to standard web browsers the XHTML documents generated from templates. An event engine handles events that occur during the use of the application, and can programatically trigger actions, thus providing a basis for the definition of dynamic visualisations.

Advene can support various video players, which however may not implement all features that are necessary to fully render hypervideos. The primary video player, VLC [2], is a versatile and cross-platform video player that supports many types of audiovisual documents on almost any medium (video file, DVD, video stream). We have created python bindings in order to be able to control it, and added some features like the overlay of SVG graphics on the video. Advene can also use the gstreamer framework[1] as video rendering system, but it will then miss the SVG overlay feature for now – we plan to develop it for gstreamer.
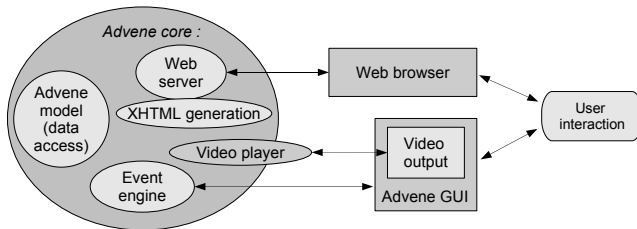


**Figure 2: The Advene software architecture**

The Advene framework is intended to foster new uses of audiovisual metadata. Its templating system and its rule-based dynamic view definitions, which will be described further, allow to quickly specify new visualisations for metadata. More elaborate tasks or visualisations can be programmed in python, the implementation language of the

---

[1]http://gstreamer.freedesktop.org/

Advene framework, and integrated in Advene through a plugin infrastructure.

## 3.  ANNOTATION CREATION AND EVOLUTION

Annotations can be created in multiple ways. Data may be imported from external applications such as feature extraction programs. Note that the Advene framework only integrates timestamped data from external sources, and is completely content-agnostic. Through the plugin infrastructure, new data importers can be defined and proposed to the user. It is then easy to integrate and test new algorithms for feature extraction or automatic processing, benefiting from the visualisation facilities of Advene.

A semi-automatic process can involve assistants that will guide the user in the creation of metadata, for instance by automatically proposing metadata based on content analysis.

Eventually, the annotations can be completely manually created, using for instance the note-taking feature of the application. The note-taking feature provides a simple means to enter time-aligned text, taking quick notes while watching the audiovisual document. Once the notes are complete, they can be converted to annotations.

Once created by any means, annotations can be refined by changing either their content, their type (transmuting a generic annotation to a more specific one...) or their timecodes (aligning on other annotations, adjusting...).

## 4.  METADATA VISUALISATION

Metadata is created to enrich the audiovisual documents with supplementary information, that can be used to more efficiently navigate or search the audiovisual documents. However, getting the relevant information is only part of the search process. The other part deals with information visualisation, which is dependant on the nature of the intended task. It is thus important to provide different visualisation means for information, that the user can customize to adapt to his needs. The current version of Advene implements three types of view: adhoc view, static views and dynamic views.

### 4.1  Adhoc views

Adhoc views are customizable GUI components, provided by the application. They offer more interaction facilities than other views, but less flexibility in their variations and user-specifiability. Some views are commonly found in audiovisual applications, like the timeline view that presents annotations on a temporal scale. Other are more specific to Advene, such as the transcription view that presents a text composed with the concatenation of a set of annotations, whose display is synchronized with the playing of the audiovisual document: it is possible to navigate from the text to the corresponding point in the audiovisual document, and the text corresponding to the current player position is also continuously highlighted.

Figure 3 presents the main Advene interface, with a timeline view at the bottom, a transcription view presenting the text panels at the right-hand side of the window, and the result of a search on the contents of annotations at the right of the video.
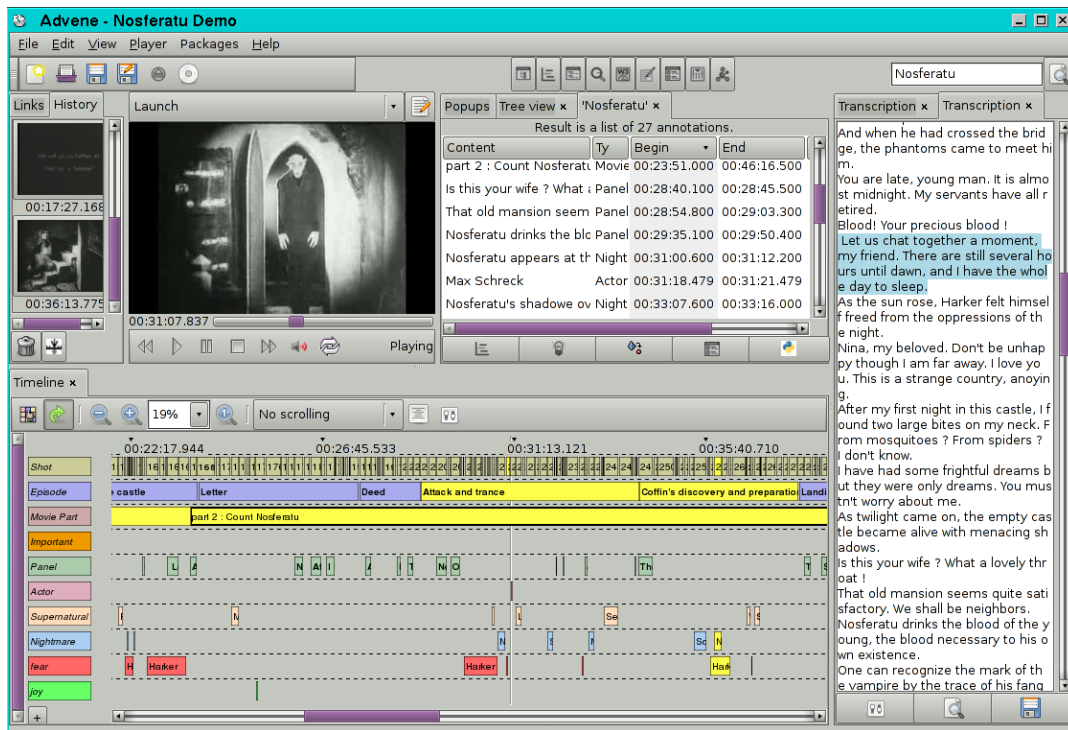
Figure 3: Advene main interface

## 4.2 Static views

Another way of visualising information is through dynamically generated hypertexts in XHTML format, using the annotation contents as well as screenshots extracted from the audiovisual document. For that purpose, the Advene prototype embeds the Zope Page Templates templating engine. It defines two entities: an addressing scheme named TALES (*Template Attribute Language Expression Syntax*) and a templating language called TAL. Both offer interesting properties. TALES presents a simple addressing scheme, using a path-like syntax, accessing the data as a simple structure. TAL relies on TALES, together with seven simple processing directives stored as XML attributes in a specific namespace. TAL templates are thus XML-valid and may be edited with any XML editor in a straightforward way. We have for instance integrated into Advene FCKEditor[2], a WYSIWYG browser-based XHTML editor in order to ease the definition of new templates.

As various Advene actions (player control, application control...) are accessible through the web server, it is possible to interact with the application from a generated XHTML document, and for instance to activate a dynamic view or an adhoc view from a static view.

A typical static view for example is a table of contents, generated from annotations, with illustrating screenshots from the video taken on-the-fly (for copyright reasons, no screenshot from the audiovisual document is stored in the package), featuring direct links to various parts of the audiovisual document. The Advene templates allow to easily design this kind of view, as well as to quickly experiment with new ideas.

## 4.3 Dynamic views

Dynamic views allow to execute various actions depending on events occurring during the audiovisual document play. They also rely on TALES to specify elements from the model. Their rule-based syntax (Event-Condition-Action)[3], inspired by classical interfaces (mail filters, etc.) aims at being accessible to end-users. Events match user interaction occurrences and metadata-related events (begin or end of annotations, etc). Conditions include notably Allen relations on fragments (contains, before, overlaps, etc.), as well as many tests on metadata contents. Actions allow to control either the video player (play, pause, go to another position, take screenshots, display text or SVG captions on the video, etc.) or elements from the GUI (display informational and navigation popups, open an URL in a web browser, etc.).

Using dynamic views, it is possible to specify simple or complex behaviours. A typical example is the captioning of subtitles over the video, which is expressed by the rule "*When the beginning of an annotation is reached (*Event*), if the annotation is a subtitle (*Condition*), then display the annotation content over the video (*Action*)*". Using player control actions allows to dynamically modify the playing of the audiovisual document. Playing a PG-rated version of a movie, given that all rated scenes are annotated, is expressed by the rule: "*When the beginning of an annotation is reached, if the annotation is a rated scene, then go directly to the end of the annotation*". Eventually, more elaborate actions allow to provide interactive navigation through the use of choice dialogs proposing to navigate from one point to another. A number of actions are provided with the Advene framework, and it is possible to implement new ones in order to augment the dynamic view possibilities.

---

[2]http://www.fckeditor.net/

## 5.  DATA EXCHANGE

Data exchange is a fundamental aspect of Advene. An Advene package is a file that contains both the annotation data and the definition of its visualisations. It is thus possible to exchange data and visualisations in order to conjointly work on the same audiovisual document and produce analyses as well as raw data (annotations) reusable by others. The independence, storage-wise, of the package from the audiovisual document means that the distribution restrictions that may apply to the audiovisual document do not apply to the package.

There are two ways of exchanging data: by copying data in order to modify it, or by importing existing data. The import mechanism only defines a reference to data (annotation, views, schemas) stored in another package, in order to benefit from its evolution.

Being single files, packages can be stored on a standalone, disconnected computer for private work, sent by e-mail or stored on web servers (as opposed to some audiovisual metadata systems that rely on databases).

## 6.  CONTRIBUTIONS

The Advene project aims at studying and creating innovative interaction modalities with audiovisual documents, by facilitating user-innovation through metadata exchange. Its metadata-centric approach focuses on metadata exchange and visualisation, and stresses the importance of its availability to the end-user. Moreover, compared to other approaches that focus either on metadata storage (as MPEG7 [4], Annodex[3]) or on visualisation (as SMIL[4]), we consider that both are interdependent and should be considered together. That is why the definition of Advene packages includes both metadata and visualisation specifications.

Eventually, thanks to the versatility in terms of supported formats (DVD, digital video files, streaming, ...) of the video player, the Advene prototype can be used to work on diverse data sources. It is especially interesting to consider DVDs, that provide a large, common and various corpus, available to end-users. Most online video systems sharing some similar features impose the conversion of the video to a unique format (Flash video), and require that the user uploads its documents, stores its annotations on their server and always has internet connectivity. With Advene, users can work offline (on a journey for instance), work with video in original formats without any conversion, and most of all, keep the control of their data (both audiovisual and annotations).

## 7.  CURRENT USES AND FUTURE WORKS

In the context of the Cinelab project, that aims at providing tools for movie critics, the Advene prototype is being used to model usages and description schemes for movie critics, as well as to define description and annotation models. It is also used in a research context by interaction researchers, in order to analyse corpuses of videos. We are also experimenting with it for practical work within a course on document annotation.

Many developments are planned for the prototype in the near or less near future. The model library, which provides an abstraction over the data for convenient access by the application and its extensions, is being rewritten to improve its performances as well as its ability to annotate concurrently multiple videos.

We are also developing the overall usability of the GUI, through a rapid test cycle made possible thanks to the python implementation. More particularly, we will improve the integration of the FCK WYSIWYG editor, and extend it to allow a more easy edition of document templates. After a few years of experimentations, we have identified common practices in the conception of static and dynamic views. Based on these common practices, GUI wizards will be generalized to assist the users in the definition of new views.

Eventually, we want to provide more automatic or semi-automatic tools to users, by integrating algorithms or applications for shot or audio segmentation, speech recognition, text recognition, etc. We are thus looking for open-source code for feature extraction that could offer useful results.

## 8.  CONCLUSION

The open-source Advene prototype offers a framework for integrating and visualising audiovisual metadata. It allows users to define by themselves, according to their current tasks, the structure of the metadata as well as the different ways that it should be visualised. Through these features, we hope to foster innovation and see new interaction modalities with audiovisual documents emerge.

## 9.  REFERENCES

[1] O. Aubert and Y. Prié. Advene: active reading through hypervideo. In *ACM Hypertext'05*, pages 235–244, Salzburg, Austria, Sep 2005.

[2] H. Fallon, A. de Lattre, J. Bilien, A. Daoud, M. Gautier, and C. Stenac. *VLC User Guide*. VideoLAN Project, 2003.

[3] N. W. Paton, editor. *Active Rules in Database Systems*. Springer Verlag, New York, 1999.

[4] J. M. M. Sanchez, R. Koenen, and F. Pereira. MPEG-7: The Generic Multimedia Content Description Standard, Part 1. *IEEE Multimedia Journal*, 9(2):78–87, Apr-Jun 2002.

---

[3]http://www.annodex.net/
[4]http://www.w3.org/TR/smil20/