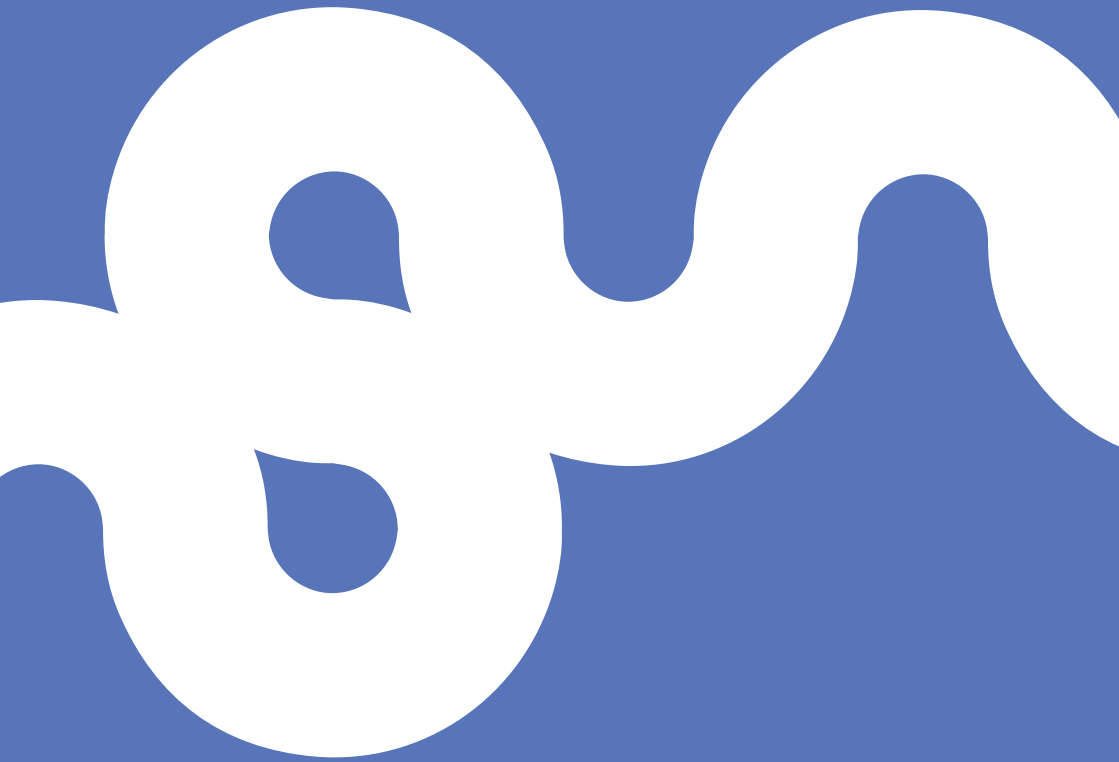
Vers une interopérabilité des annotations vidéo ?

Olivier Aubert

Vers une interopérabilité des annotations vidéo ?



OLIVIER AUBERT

Introduction

Au cours des deux dernières décennies, les pratiques d'annotation vidéo se sont considérablement développées dans les milieux de la recherche, de l'enseignement et de la création artistique. Elles répondent à une nécessité commune : rendre les contenus audiovisuels explorables, analysables et partageables grâce à des dispositifs permettant de décrire, d'enrichir et de structurer les données temporelles. Or, cette diversité de pratiques s'est accompagnée d'une prolifération d'outils et de formats d'annotation, souvent incompatibles entre eux. Cette hétérogénéité est facteur de diversité mais complique parfois la mutualisation des corpus, la reproductibilité des analyses et la pérennisation des données.

Dans ce contexte, la question de l'interopérabilité apparaît comme un enjeu scientifique et technique central. Elle désigne la capacité d'un système à dialoguer efficacement avec un autre, en assurant une compréhension mutuelle des données échangées. Ce document s'inscrit dans la continuité du European Interoperability Framework (EIF), qui définit les fondements de cette notion sous plusieurs dimensions - politiques, organisationnelles, légales, techniques, syntaxiques et sémantiques. Nos travaux se concentrent ici sur les trois dernières, plus directement opérationnelles.

L'objectif de ce document est double. D'une part, il vise à fournir un aperçu des formats et modèles existants, qu'ils soient standardisés (par exemple IIIF, WebAnnotation, ou REFI-QDA) ou spécifiques à certains outils (ELAN, Advene, Celluloid, BORIS...). D'autre part, il s'efforce d'identifier les axes de convergence possibles entre ces approches, afin de proposer des formats d'interopérabilité adaptés aux besoins des communautés de pratique. Cette démarche s'est entre autres nourrie de scénarios d'usages concrets développés lors des ateliers CANEVAS organisés en mai 2025 et d'échanges avec des praticiens. Cette

contribution vise ainsi à la fois à constituer un outil d'analyse critique des solutions existantes et à proposer des pistes pour la conception de formats et de procédures d'échange normalisés.

Contexte général de l'interopérabilité

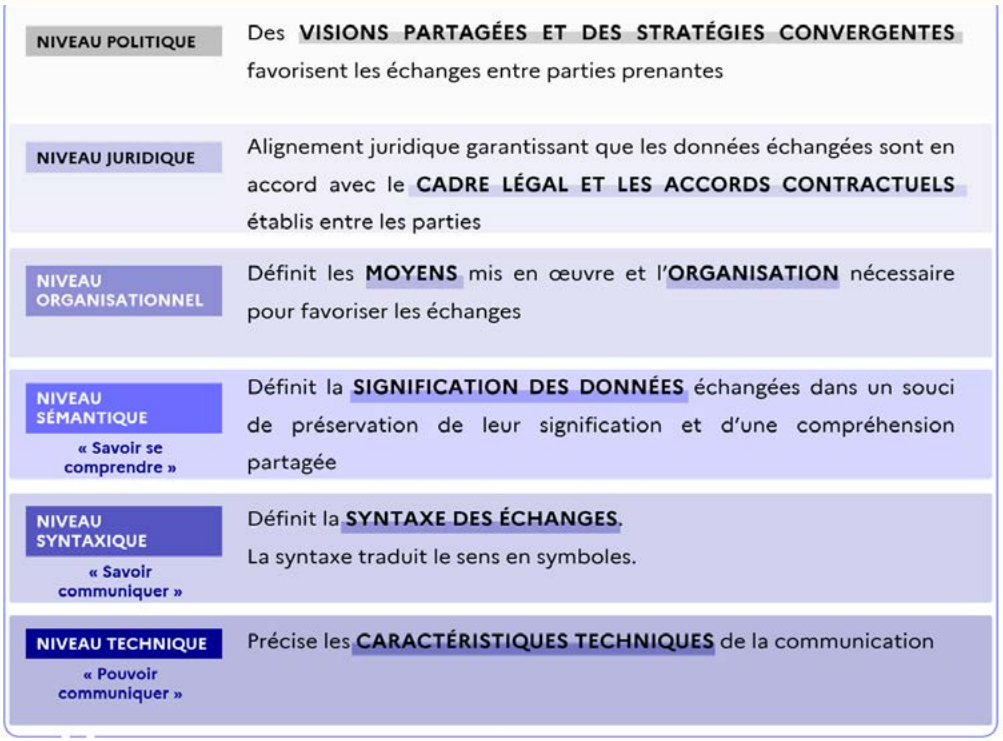
Le Référentiel Général d'Interopérabilité (RGI Version 2.0, Direction Interministérielle des Systèmes d'Information et de Communication, transcription nationale de l'European Interoperability Framework - EIF) définit six niveaux d'interopérabilité. Dans le présent document, nous ne considérerons pas - hormis en élément de contextualisation - les 3 niveaux les plus élevés : interopérabilité politique, interopérabilité légale et l'interopérabilité organisationnelle. Nous nous attacherons aux questions de l'interopérabilité technique « pouvoir communiquer », l'interopérabilité syntaxique « savoir communiquer » et l'interopérabilité sémantique « savoir se comprendre ».

Comme le précise le Référentiel d'interopérabilité des services numériques pour l'éducation :

- Le niveau technique définit les caractéristiques techniques de l'échange comme les protocoles et moyens de transport.

- Le niveau syntaxique précise le format des données échangées.

- Le niveau sémantique permet une compréhension partagée des différents éléments de l'échange entre les tiers concernés. Elle pourra ainsi reposer sur un ensemble de nomenclatures partagées entre les acteurs.



Définition et niveaux d'interopérabilité [RISNE]

Annotation vidéo - contexte et définitions

Le développement des outils d'annotation vidéo s'est effectué conjointement avec les possibilités techniques de visualisation et manipulation de la vidéo. Le codec MPEG2 (utilisé par les DVD) est apparu en 1994, le codec MPEG4 (et ses dérivés/variantes populaires telles que DivX) en 1999. Ces développements ont rendu possible l'instrumentation de lecteurs vidéo offrant un accès aléatoire aux différents moments des vidéos [HyperCafe1996].

Les outils d'annotation vidéo ont principalement émergé à partir des années 2000 avec le développement des possibilités techniques de la vidéo numérique, pour répondre aux besoins de communautés particulières, notamment en recherche. Ainsi le logiciel Anvil a été créé en 2000 pour outiller la pratique de chercheurs en analyse comportementale. Transana a été créé en 2001 afin de permettre le codage de corpus vidéo dans le domaine pédagogique ou d'études sociologiques. Advene a été créé en 2002 dans l'objectif d'étudier les interfaces de manipulation des annotations et de création d'hypervidéos, visant initialement les champs de l'analyse critique, puis s'étendant à des champs divers (arts, sociologie, etc). ELAN a été créé en 2002 également, par un laboratoire de linguistique, afin d'instrumenter ses pratiques d'analyse du langage et des postures.

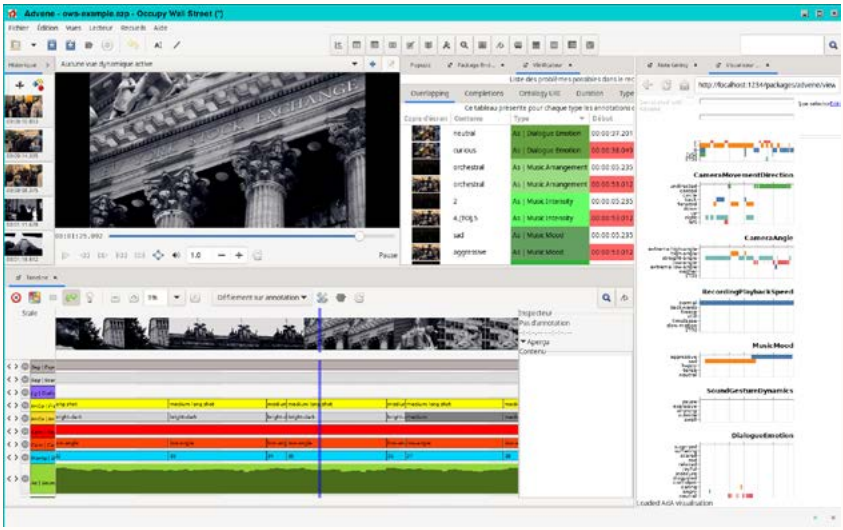
Ces logiciels sont tous des applications natives (à télécharger et installer sur un poste de travail), la vidéo n'étant pas encore bien gérée par les navigateurs web. Le logiciel Lignes de Temps, développé à partir de 2006 à l'IRI et qui visait les pratiques de critiques de film professionnelles et amateur, a fait le choix d'une interface web via un plugin Flash, qui était nécessaire pour visualiser et interagir avec la vidéo au sein des navigateurs web. La technologie propriétaire Flash a posé de nombreux problèmes notamment de

sécurité, et les standards du web ont évolué dans les années 2010, rendant ainsi possible l'interaction avec les vidéos sans intermédiaire au sein des navigateurs web. La technologie Flash a été désactivée en 2020, entraînant la non-disponibilité du logiciel Lignes de temps.

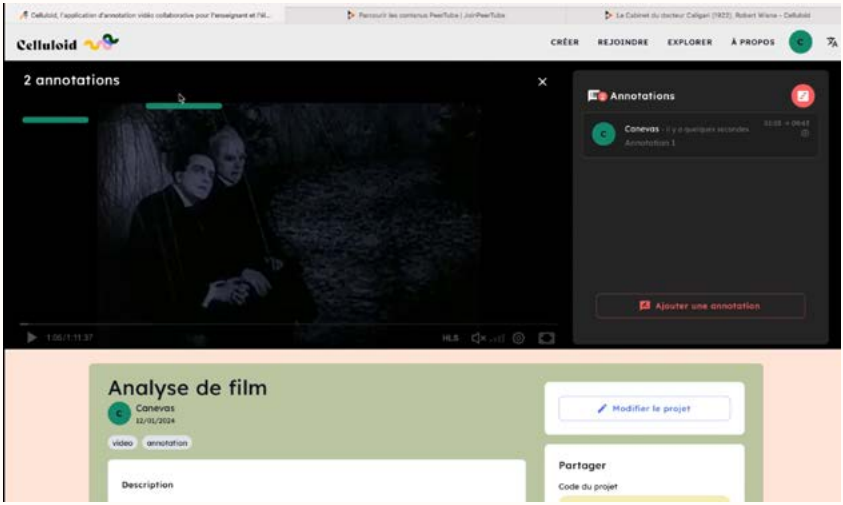
L'évolution des technologies liées à la vidéo, notamment dans l'écosystème web mais aussi dans différents frameworks, a conduit à une explosion d'usages dans plusieurs domaines : linguistique (Anvil, ELAN, Exmeralda, DOTE...), analyse de média (Dicto, AV-Annotate...), pédagogie (Celluloid, COCoNotes, Mediathread...), performances artistiques (MotionBank, MemoRekall, Arvest...), ethologie (BORIS, Simba...), Sport (LongoMatch, CoachLogic...), informel (YiNote, Reclipped...).

Au cours des années 2010, la progression des algorithmes d'intelligence artificielle a entraîné un besoin énorme de données d'annotation (y compris vidéo) pour entraîner les modèles, d'où le développement de plusieurs outils d'annotation/labellisation (Deepen.ai, Vatic, LabelMe...) souvent orientés annotation graphique d'éléments vidéo et optimisés pour l'annotation en masse.

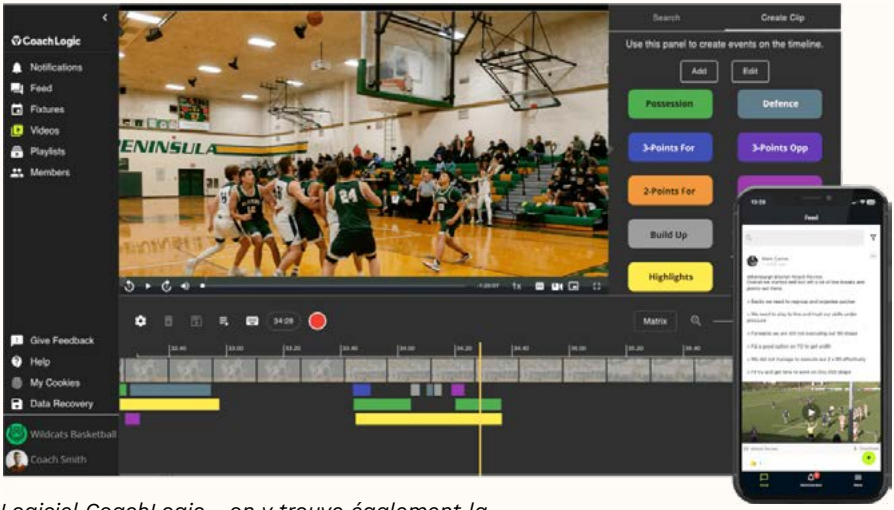
Chacune de ces communautés a orienté les caractéristiques des différents logiciels.



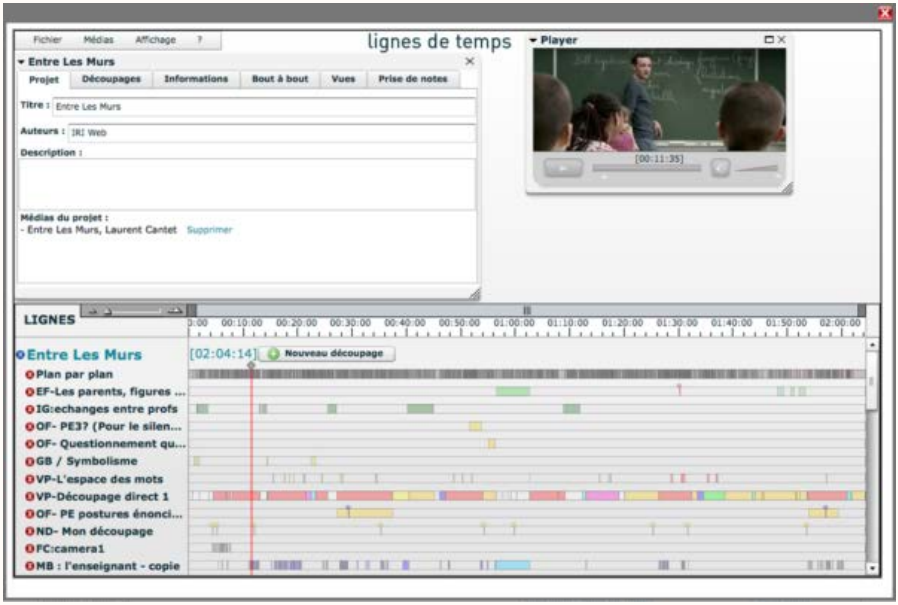
Logiciel Advene - on voit les différents types d'annotation présentés sous forme de lignes.



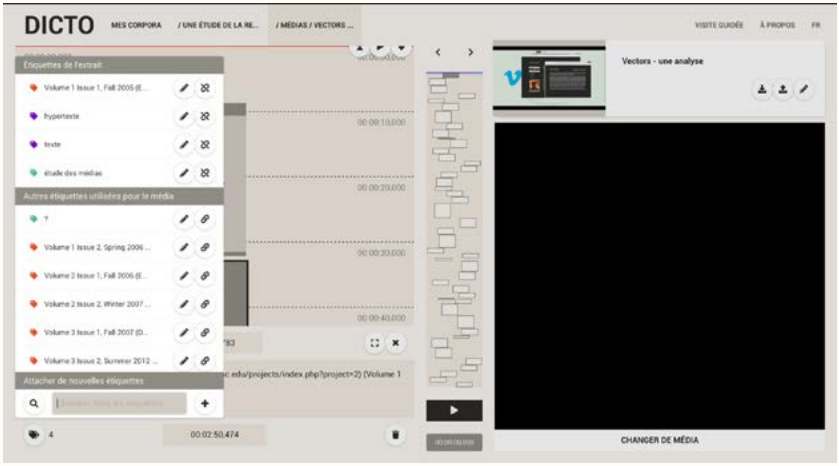
Logiciel Celluloid - l'accent est mis sur la vision de l'ensemble des annotations.



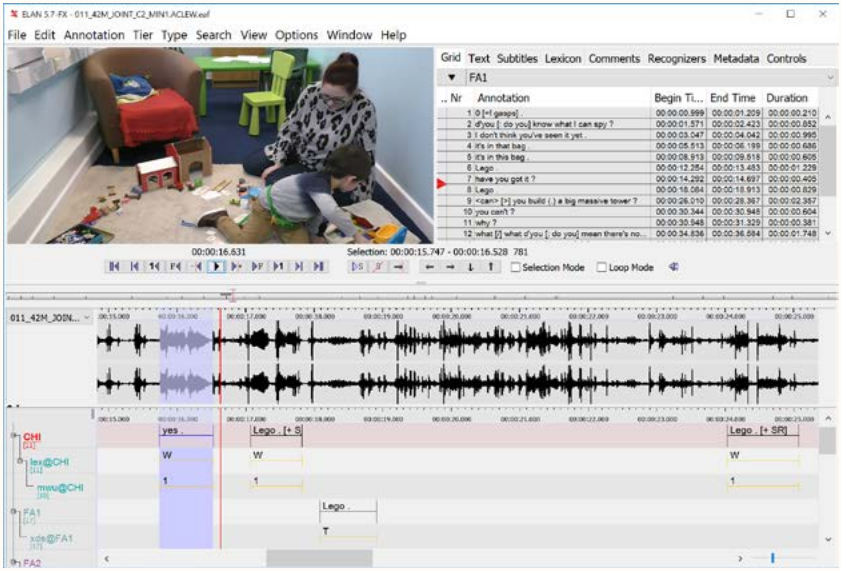
Logiciel CoachLogic - on y trouve également la possibilité de catégoriser les annotations.



Logiciel Lignes de temps : il était centré sur la métaphore des lignes d'annotation et de leur manipulation.

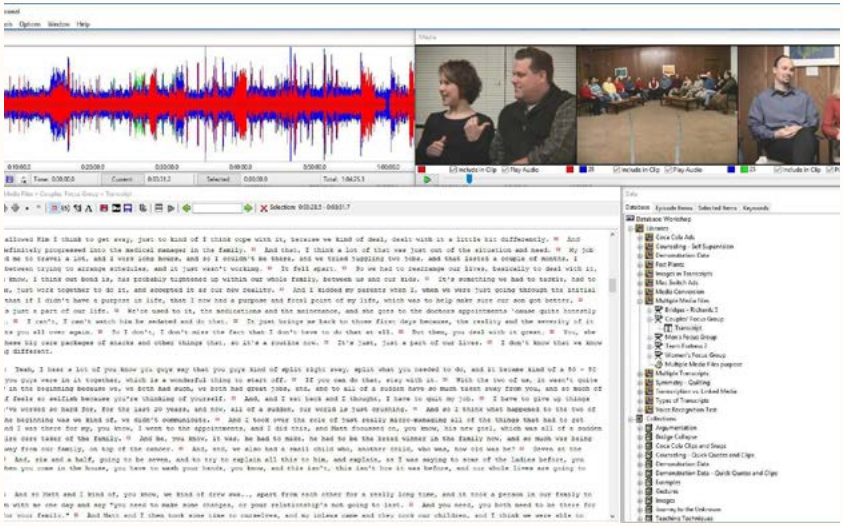


Logiciel Dicto - il présente le temps sous forme verticale.

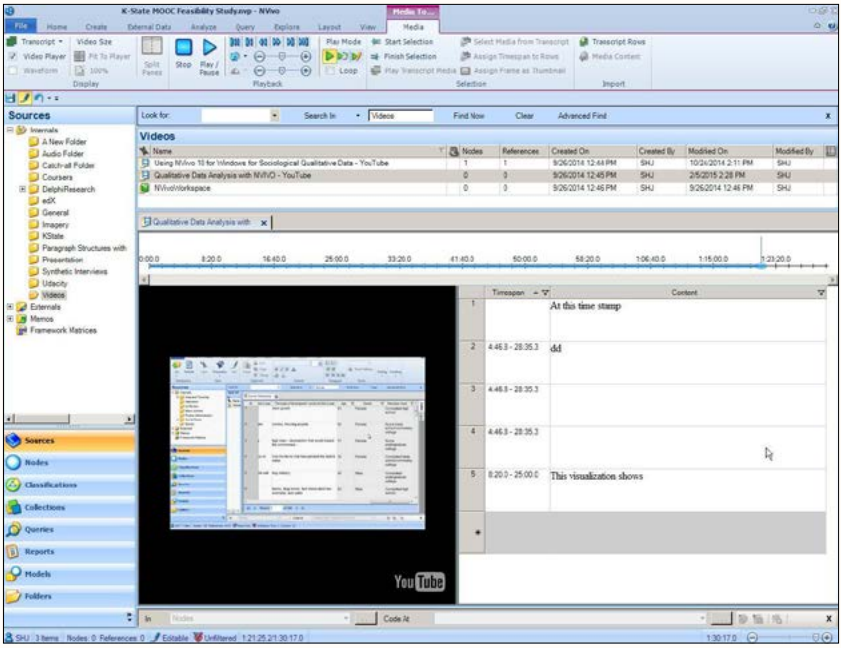


Logiciel ELAN - on peut voir le caractère hiérarchique des Tiers (groupements d'annotations).

Logiciel Mediathread - à visée pédagogique, il vise à encourager les échanges (feedback) autour des annotations.



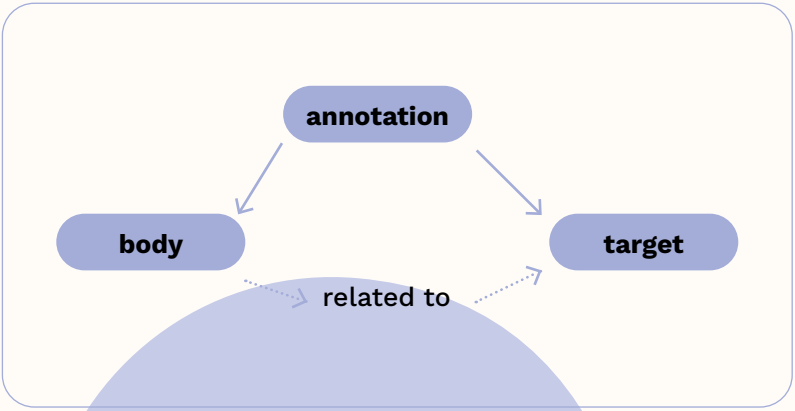
Logiciel NVivo - un des acteurs importants des logiciels d'analyse qualitative (CAQDAS).



Logiciel Transana - un autre représentant des CAQDAS.

Modèle de référence : WebAnnotation

Le consortium W3C a produit en 2017 une *Technical Recommandation* [WebAnnotation] définissant un modèle de données, un vocabulaire et des protocoles associés concernant l'annotation (en général) sur le web. Le modèle WebAnnotation est maintenant le cadre communément utilisé pour dénommer les éléments d'une annotation : une annotation met en relation une cible (*target*) avec un corps (*body*) qui correspond au contenu de l'annotation.



Modèle de données WebAnnotation [WebAnnotation].

Dans le cas des annotations vidéo, la cible (target) peut être définie à travers une URI de type MediaFragment [MediaFragment] encodant dans l'URI les paramètres d'ancre temporelle, par exemple http://www.example.com/example.ogv#t=10.434,20.200 ou via un sélecteur (FragmentSelector) qui permet de séparer l'URL de la source (http://www.example.com/example.ogv dans l'exemple précédent) du fragment lui-même (#t=10.434,20.200). Le standard WebAnnotation définit en outre la notion de *AnnotationCollection* qui permet de constituer et désigner des ensembles cohérents d'annotations, de manière non exclusive (une annotation peut faire partie de différentes collections).

Dimensions d'interopérabilité

Dans le domaine de l'annotation vidéo, et au delà des problèmes classiques liés à la vidéo comme les paramètres et formats d'encodage vidéo, ainsi que la désignation précise de la vidéo concernée et de ses modalités d'accès, il y a principalement deux dimensions à considérer :

- d'une part la **dimension spatio-temporelle**, c'est-à-dire l'expression du lien avec le fragment vidéo (le sélecteur dans le modèle WebAnnotation). La très grande majorité des logiciels d'annotation vidéo considère une portée temporelle définie explicitement, avec un début et une fin. Il existe cependant certains usages où seule une borne de début est spécifiée, souvent à des fins de simplification [VideoAnt], par exemple dans la fonctionnalité de signet, ou de repères dans des notes horodatées. De plus, selon les scénarios d'usage et les logiciels utilisés, les questions de résolution temporelle peuvent être abordées différemment, avec des impacts sur d'autres dimensions telles que le nombre d'annotations considérées : de

nombreuses annotations couvrent des segments de plusieurs secondes, mais le développement des outils de transcription automatique et des interfaces de visualisation associées a conduit également au développement de transcription (donc d'annotations) au mot près, ce qui produit un grand nombre d'annotations. On retrouve cela également en linguistique ou analyse comportementale, où l'on peut avoir besoin d'une précision de l'ordre de la milliseconde. Enfin, les annotations vidéo étaient pour la plupart originellement liées uniquement à un fragment temporel, mais les usages se développent dans l'annotation d'objets/zones (statiques ou dynamiques) à l'intérieur d'une vidéo.

- d'autre part la **dimension typage/codage/structuration des contenus d'annotation** qui relève de l'interopérabilité sémantique évoquée plus haut. Les contenus d'annotation peuvent être en texte libre, ou dans des formats semi-structurés (tels que HTML ou Markdown), voire dans des formats structurés (JSON, XML...) pour contenir des informations multivaluées ou structurées. On rejoint ici le domaine des logiciels d'analyse qualitative (CAQDAS) et les enjeux de codage. De plus les annotations elles-même peuvent être typées (catégorisées) afin d'en faciliter l'exploitation. Ainsi, les logiciels orientés linguistique tels que ELAN, Anvil, PRAAT utilisent la notion de Tier pour catégoriser les annotations. On retrouve une notion similaire sous l'appellation de *Type d'annotation* dans Advene, ou de lignes dans Lignes et Temps.

Il peut y avoir de plus des dimensions qui ne relèvent pas de l'annotation vidéo elle-même, mais plutôt de son processus d'utilisation. Ainsi, dans un espace collaboratif, la notion d'auctorialité est importante à transcrire et préserver (savoir qui a créé une annotation, qui l'a modifiée). Ces éléments relèvent de la dimension de **structuration des métadonnées d'annotation**.

Ces différentes dimensions sont plus ou moins intégrées dans les différents logiciels d'annotation vidéo. Il convient donc pour définir un format d'interopérabilité de bien préciser les communautés qui sont impliquées ainsi que les usages envisagés.

Travaux antérieurs/connexes

La problématique de l'interopérabilité n'est pas récente. Plusieurs travaux ont déjà abordé ces questions, sans forcément de réponse définitive.

Au niveau international, suite à des discussions au sein du groupe de travail Digital Annotation de DARIAH et le groupe AVinDH (Audiovisuel dans les Humanités Numériques) de l'ADHO, l'infrastructure de recherche CLARIAH a organisé en 2020/2021 un groupe de travail VAINT (Video Annotation Interoperability) pour avancer sur les questions d'interopérabilité des logiciels d'annotation vidéo avec les créateurs des logiciels ELAN, Web Annotation Tool, Video Annotation Tool, Frametrail et Advene. Le groupe de travail n'a pas abouti à la production d'une référence finale, mais a tout de même convergé vers l'utilisation du modèle WebAnnotation (inspiré en grande partie par IIIF), qui a été implémenté par plusieurs logiciels dont ELAN et Advene. Il a proposé également un schéma positionnant les concepts et processus impliqués.

En France, plusieurs groupes ont également abordé cette question, directement ou indirectement. Le réseau thématique TIPS-IA de MATE-SHS est en train de travailler sur les transcriptions et les traitements sur les données de transcription (pseudonymisation, analyse...) et l'interopérabilité des outils est un des aspects. Dans la communauté de recherche autour des sciences du langage, le consortium HumaNum CORLI (CORpus, Langues et Interactions) anime un groupe de travail sur l'interopérabilité https://corli.huma-num.fr/fr/les-groupes-reseaux/gp1/ dans le contexte des corpus oraux ou multimodaux, qui sont

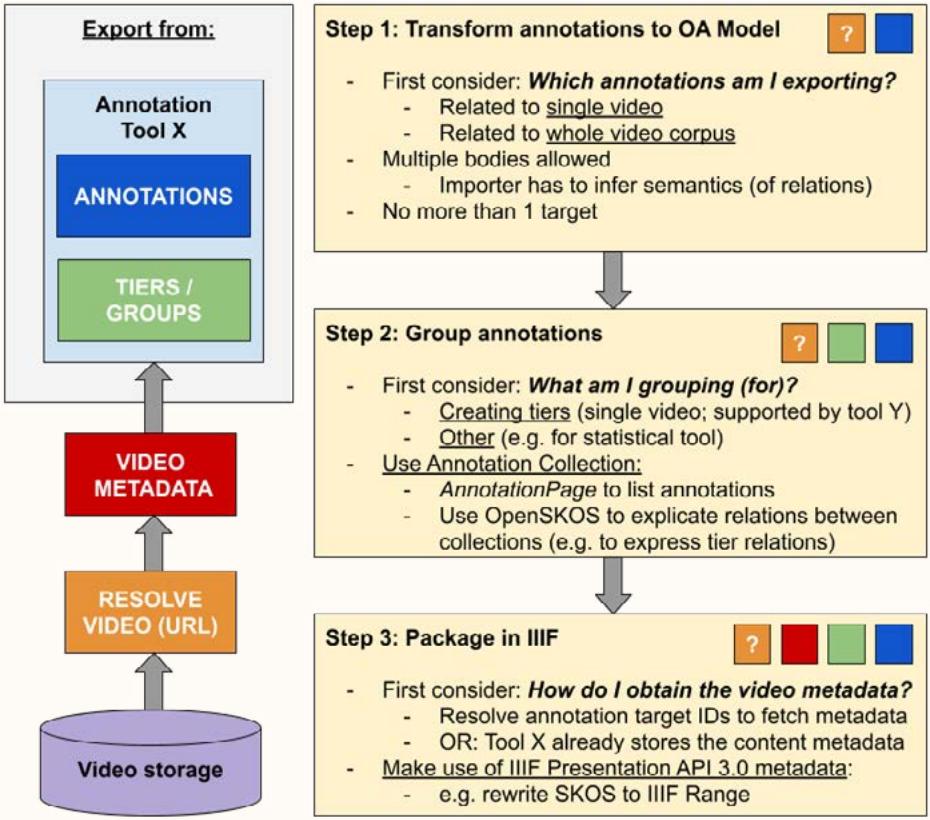
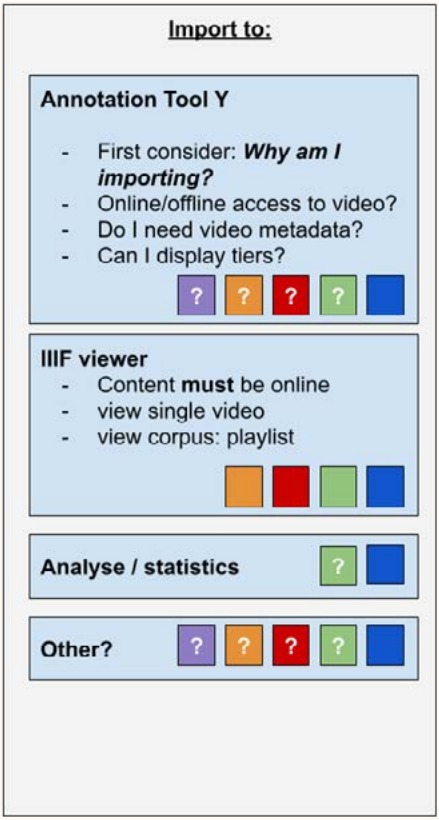


Schéma de principe des processus impliquant l'interopérabilité (Jaap Blom, VAINT working group).

une des communautés de pratique autour de l'annotation de médias temporalisés. Ils préconisent dans leur FAQ l'utilisation de formats utilisables par les outils du champ (PRAAT, CLAN…) ou de TEI (format structuré d'édition de texte). L'outil TEICorpo permet la conversion entre ces différents formats. Étant centrés sur le langage, ces formats mettent l'accent sur la dimension textuelle/transcription, ce qui peut parfois limiter la généralisation à des pratiques plus larges d'annotation vidéo. Il en va de même pour les outils recensés par Thibaut Rioufreyt [Rioufreyt 2018]



dans son article “La transcription outillée en SHS. Un panorama des logiciels de transcription audio/vidéo” qui aborde la question de l'interopérabilité, en se focalisant sur l'aspect transcription.

Synthèse des ateliers CANEVAS organisés en mai 2025

Des ateliers en ligne ont été menés pour recueillir une expérience de terrain ainsi que des attentes dans le domaine de l'interopérabilité. Dans la communauté Celluloid, il n'y a pas de scénario d'interopérabilité vraiment établi, mais des besoins d'exporter les informations (tableur, etc) pour traitement ultérieur, et des besoins en fonctionnalités (catégorisation, etc).

Concernant les pratiques d'interopérabilité, un des formats naturels qui s'impose pour conserver les timecodes est le format **.srt / .vtt**, qui a l'avantage d'un support très répandu, notamment des lecteurs vidéo standard. Mais il va être mobilisé pour assurer une interopérabilité ciblée sur des usages particuliers, où on va exporter précisément un type d'information, pour interopérer avec un outil spécialisé ou pour visualiser simplement une information basique dans un lecteur vidéo standard.

En interopérabilité plus générale, **IIIF** a été évoqué et il présente effectivement l'avantage d'une normalisation et d'une communauté active, avec la limite néanmoins que IIIF est un format de présentation (du moins le standard le plus identifié, Presentation), et non un format pensé pour l'échange d'annotations plus général (sans forcément de présentation imposée derrière). IIIF est utilisé nativement par les logiciels Arvest et AV-Annotate. Cependant, il mobilise en sous-jacent le modèle (standard) **WebAnnotation**, avec quelques contraintes sur la structure (et la communauté IIIF réfléchit à ajouter quelques autres contraintes). WebAnnotation est aussi la solution préconisée par le groupe CLARIAH VAINT qui s'était tenu entre 2018 et 2021 et qui a été implémentée notamment par Advene et ELAN.

Scénarios d'interopérabilité

Différents scénarios d'interopérabilité, existants ou envisagés, ont été identifiés au cours des ateliers.

Scénario 1

Transcription de vidéo avec Whisper puis intégration des annotations ainsi produites dans Advene/ELAN/ Celluloid

Dans ce scénario, on souhaite intégrer dans un outil cible (Celluloid par exemple) les annotations issues d'un traitement automatique tel que la transcription automatique. Les informations importantes sont le contenu textuel (la transcription) et les bornes temporelles. Les autres informations (auctorialité, structure) ne sont pas nécessaires. En pratique, on utilise souvent le format **srt** comme format de sortie du logiciel de transcription, qui peut être importé par les différents logiciels.

Scénario 2

Saisie d'annotations collaboratives dans Celluloid, puis utilisation dans un autre logiciel tel qu'Advene pour poursuivre l'analyse et l'exploitation

Une communauté de chercheurs utilise Celluloid pour annoter collaborativement une vidéo. Pour poursuivre leur analyse, ils ont besoin de catégoriser les annotations produites. Ils se tournent alors vers un autre logiciel tel qu'Advene pour reprendre l'ensemble des annotations produites et les catégoriser au travers de différents types d'annotations. Il souhaitent préserver l'information d'auctorialité des annotations. Ils peuvent alors utiliser le format **csv** d'export des annotations, qui préserve ces informations, et l'importer au sein d'un logiciel permettant la catégorisation pour poursuivre le travail d'analyse.

Scénario 3

Dépôt des annotations produites par un logiciels sur un dépôt type Nakala/Zenodo pour préservation

Afin d'assurer la pérennité des informations, Nakala préconise d'utiliser des formats ouverts et propose de suivre les recommandations du CINES. Dans les formats proposés sur https://facile.cines.fr/ le format CSV (assimilé à texte encodé en UTF-8) répond aux besoins et peut être envisagé pour préserver une bonne partie des informations d'annotation (contenu, ancres, structure, auctorialité). La plupart des logiciels d'annotation vidéo proposent un export dans ce format permettant de transcrire les différentes dimensions des annotations vidéo (temporelle, contenu, catégorisation, auteur), et permettent également la plupart du temps d'importer des données depuis ce format (avec des possibilités de personnalisation variables).

Formats d'interopérabilité

Parmi l'ensemble des formats cités, nous avons sélectionné quelques uns des formats d'interopérabilité existants, avec comme critères impératifs ou optionnels :

- être utilisable (en import et/ou export) par au moins 2 outils
- pouvoir exprimer l'information temporelle
- pouvoir représenter un contenu textuel simple
- Pouvoir optionnellement représenter un contenu structuré (catégories, types, etc)
- pouvoir adresser plusieurs vidéos
- avoir la capacité de préserver les métadonnées d'auteur

Certains formats répondaient à plusieurs critères mais n'ont pas été retenus dans la sélection pour à cause d'une moins grande généralisation possible. Ainsi les formats centrés texte tels que TextGrid (originaire de PRAAT, utilisé par de nombreux logiciels dans la communauté de traitement du langage) ou TEI, qu'on peut retrouver dans [Rioufreyt 2018] ne semblent pas forcément adaptés pour un usage d'annotation plus général. Il a semblé plus pertinent de cibler sur les formats les plus génériques et les plus opérationnalisables possibles.

Critère	WebAnnotation (JSON-LD)	REFI-QDA (XML)	CSV / TSV	srt
Portée temporelle	Début/Fin (FragmentSelector)	Plage/Point temporel (dans les code-segments)	Début/Fin (colonnes distinctes)	Début/Fin
Résolution (ms)	Oui	Oui	Possible	Oui
Annotation Spatiale	Oui (Selector: SVG/ Box)	Non (axé analyse qualitative)	Non	Non
Contenu catégorisé (Typage/ Tier)	Oui (via motivation ou body en JSON-LD)	Oui (Modèle de Codage/ Catégorisation)	Possible (une seule colonne de catégorie par ligne)	Non
Contenu structuré	Oui (JSON-LD)	Oui (Modèle de Codage/ Catégorisation)	Possible (via multiples colonnes)	Non
Métadonnées d'Auctorialité	Oui (via creator, modified, etc.)	Oui (via meta-information au niveau du projet)	Possible (ajout de colonnes dédiées)	Non
Support de multiples vidéos	Oui (utilisation de sources différentes)	Oui	Possible (ajout de colonnes dédiées)	Non
Export possible par	Arvest, AVAnnotate, Advene, ELAN	NVivo, Atlas.ti, Transana	Advene, ELAN, Celluloid, BORIS, Transana...	Advene, ELAN, Celluloid, Transana...
Import possible par	Arvest, AVAnnotate, Advene, ELAN	NVivo, Atlas.ti, Transana	Advene, ELAN, BORIS, Transana...	Advene, ELAN, BORIS, Transana...
Facilité d'implémentation/ export	Modérée (format riche, aux multiples possibilités)	Faible (standard XML complexe)	Élevée (import/ export natif de la majorité des outils)	Élevée (import/ export natif de la majorité des outils)

Table : proposition de formats cibles génériques

WebAnnotation

Le standard WebAnnotation a été proposé en 2017 par le W3C. Il définit un modèle de données (voir plus haut) pour les annotations et une représentation au format JSON-LD. Sa richesse, son extensibilité et son utilisation des standards techniques ouverts en font le modèle et format de choix pour l'expression des annotations. Il est déjà largement utilisé pour l'annotation d'images fixes par le consortium IIIF (voir plus bas), et ce dernier travaille activement à son extension aux média audiovisuels. Le format WebAnnotation est nativement implémenté par des logiciels tels que Arvest ou AV-Annotate. D'autres outils d'annotation vidéo comme ELAN et Advene le mettent en oeuvre à travers des fonctionnalités d'export/import.

REFI-QDA

REFI-QDA est un format d'échange basé sur XML pour les données issues de logiciels d'analyse qualitatives (CAQDAS), défini en 2019. Il vise à faciliter la migration des projets d'analyse (données, codages, mémos) entre différents logiciels propriétaires d'Analyse Qualitative de Données (par exemple NVivo, Atlas.ti, Transana...). Au sein du projet CORLI, Christophe Paris a intégré ce format dans sa moulinette TEI-Corpo pour combiner du texte/transcription et des catégories.

Ce format permet d'exprimer non seulement les annotations brutes, mais également la structure du projet d'analyse qualitatif (la structure de codage).

csv/tsv

Le format de données tabulaires CSV/TSV n'est pas un format d'annotation vidéo à proprement parler : c'est un format générique permettant d'exprimer sous la forme d'un fichier texte des données catégorisées en colonnes. Il est largement utilisé comme format pivot pour l'import/export de données vers ou depuis des tableurs ou des bases de données, et préconisé dans les formats ouverts pour l'archivage.

Les fichiers CSV sont structurés en colonnes (champs)

et en lignes (enregistrements). La première ligne définit généralement l'en-tête (les noms des colonnes, par exemple *start_time, end_time, content, creator*...). Les lignes suivantes contiennent les données, les valeurs de chaque colonne étant séparées par un caractère délimiteur (la virgule pour CSV, la tabulation pour TSV).

Sa force est sa simplicité d'implémentation et sa lisibilité humaine immédiate. Il permet d'échanger des jeux de données plats (sans relations complexes), mais l'interopérabilité sémantique repose entièrement sur le fait que les entêtes de colonne soient comprises de la même manière par les différents outils. C'est un format pivot d'une grande praticité, mais d'une faible richesse syntaxique et sémantique comparativement à WebAnnotation.

srt

srt est un format permettant d'exprimer les sous-titres textuels liés à une vidéo. Conçu à l'origine pour le logiciel SubRip, il est devenu un format d'interopérabilité de facto pour un grand nombre d'applications audiovisuelles, en premier lieu des lecteurs vidéos. Il a servi de base à la définition du standard WebVTT, qui en reprend les principes de syntaxe avec quelques ajustements. WebVTT n'a pas été considéré ici, car s'il apporte des extensions (formattage, métadonnées) par rapport à srt, ces extensions sont plus ou moins bien gérées par les logiciels qui le supportent. Par sa simplicité, srt reste le dénominateur commun le plus répandu.

Il utilise une structure séquentielle très simple, répétée pour chaque sous-titre :

- Numéro de séquence (ex. 1, 2, 3...)
- Plage de temps (début -> fin, avec millisecondes, sous la forme : *HH:MM:SS,mmm -> HH:MM:SS,mmm*)
- Texte du sous-titre.

IIIF

IIIF (International Image Interoperability Framework) est un standard international pour décrire et présenter des images et des métadonnées sur les images sur le web. S'il s'est historiquement largement déployé sur des images fixes, le format intègre également la possibilité d'avoir des images animées (donc des vidéos), avec un support plus ou moins avancé dans les outils de visualisation.

IIIF définit principalement 4 API :

- Image API : définit un format d'URI pour adresser des images en précisant un niveau de zoom, une zone, une qualité, une rotation, un format, etc.

- Presentation API : définit une API d'accès et surtout un format reposant sur JSON-LD permettant de définir la structure et le mode de présentation de contenus basés sur les images (intégrant des annotations).

- Search API : définit une API permettant de rechercher des annotations dans une ressource IIIF (un manifeste, une collection, etc)

- Content State API : définit une API et un schéma JSON-LD pour transmettre des informations visant à présenter d'une manière particulière un ou des contenus IIIF (Manifestes, Collections...) - mais c'est une API récente, peu supportée par les visualiseurs

IIIF et vidéo

Il existe au sein de IIIF un groupe dédié aux vidéos : https://iiif.io/community/groups/av/ qui vise à partager les expertises, pour notamment nourrir les scénarios recensés dans le *IIIF Cookbook* https://iiif.io/api/cookbook/ Le document IIIF AV Interoperability testing recense le niveau de support des différentes fonctionnalités dans plusieurs lecteurs IIIF.

IIIF et annotations

Concernant les annotations (génériques, donc incluant les annotations faites sur des vidéos), IIIF propose des serveurs d'annotation (IIIF annotation server).

Les annotations sont structurées en AnnotationCollection qui peuvent elles-mêmes être structurées dans des AnnotationPage (voir schéma[1]). «*Annotations are made available in IIIF via Annotation Pages, where typically the included Annotations target the same resource or part of it.*»

1 https://iiif.io/api/cookbook/recipe/0309-annotation-collection/diagram-309.png

Recommandations pour l'interopérabilité au sein de CANEVAS

Il n'est pas vraiment possible de définir un format unique d'interopérabilité répondant à tous les besoins. Les angles d'attaque spécifiques de chaque communauté (le langage/texte pour l'annotation linguistique, le codage et les catégorie pour la communautés CAQDAS, etc) peuvent conduite chaque groupe à favoriser un format pivot plutôt qu'un autre. Cependant, dans une optique plus agnostique et opérationnelle, il est possible de recommander l'utilisation de deux formats simples et ouverts pour assurer une interopérabilité effective et à faible complexité des logiciels.

Pour des contextes ciblés d'échange d'informations tirées des annotations, on pourra utiliser **srt** (format de sous-titres). Ce format ne permet pas d'exprimer de dimensions au delà d'un simple contenu textuel lié à deux bornes de temps sur un unique média, mais cette simplicité le rend en pratique disponible pour la quasi totalité des logiciels.

Pour des contextes nécessitant plus d'informations comme des informations de structure (type d'annotation) ou d'auctorialité, ainsi que pour l'archivage

pérenne des informations, le format **csv/tsv** permet de sérialiser ces informations, avec quelques écueils relatifs au nommage et contenu des champs/colonnes qui n'est pas standardisé. De nombreux logiciels (ELAN, Transana) proposent des assistants d'import permettant de s'adapter aux différents nommages et contenus. Par exemple, ELAN permet d'exprimer les temps de début ou de fin d'annotation sous la forme de secondes (par exemple 12.653), de millisecondes (par exemple 12653) ou d'un timestamp formatté (par exemple 00:00:12.653).

Enfin, dans une perspective d'évolution des outils, nous pouvons recommander, à l'instar d'autres recommandations (W3C, IIIF, VAINT), l'utilisation du standard **WebAnnotation** qui dispose des bonnes propriétés en termes d'expressivité et d'adéquation à l'écosystème technique actuel. La mise en pratique de cette recommandation passera d'un côté par les usages, et de l'autre probablement par la recommandation de pratiques concernant la modélisation des informations.

Remerciements

Nous tenons à remercier l'ensemble des participants aux ateliers CANEVAS pour leurs apports, et plus particulièrement
Justine Lascar et Daniel Valero de l'équipe ICAR pour leurs contributions.

Bibliographie

[RISNE] Référentiel d'interopérabilité des services numériques pour l'éducation https://doctrine-technique-numerique.forge.apps.education.fr/interoperabilite/texte/2-principes-interoperabilite/ consulté le 03/07/2025

[RGI] Référentiel Général d'Interopérabilité https://www.numerique.gouv.fr/offre-accompagnement/reference-interoperabilite-rgi/ consulté le 03/07/2025

[Hypercafe1996] Nitin «Nick» Sawhney, David Balcom, and Ian Smith, «HyperFace: Narrative and Aesthetic Properties of Hypervideo», Hypertext '96 Proceedings, ACM, New York. pp. 1-10.

[MediaFragment] https://www.w3.org/TR/media-frags/ consulté le 01/07/2025

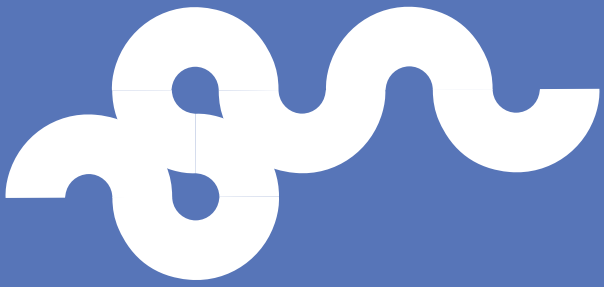
[WebAnnotation] https://www.w3.org/TR/annotation-model/ consulté le 01/07/2025

[REFI-QDA] Evers, Jeanine; Caprioli, Mauro Ugo; Nöst, Stefan & Wiedemann, Gregor (2020). What is the REFI-QDA Standard: Experimenting With the Transfer of Analyzed Research Projects Between QDA Software. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 21(2), Art. 22, http://dx.doi.org/10.17169/fqs-21.2.3439

[Rioufreyt 2018] Thibaut Rioufreyt, « La transcription outillée en SHS. Un panorama des logiciels de transcription audio/vidéo « *Bulletin of Sociological Methodology/ Bulletin de Méthodologie Sociologique*, n° 139, https://doi.org/10.1177/0759106318762455.

RÉSULTAT DU GROUPE DE TRAVAIL
DU CONSORTIUM HUMA-NUM CANEVAS

Comment faire dialoguer les outils d'annotation vidéo ? Ce groupe de travail du Consortium Huma-Num Canevas cartographie les enjeux techniques et sémantiques de l'interopérabilité. Des logiciels spécialisés (ELAN, Advene, Celluloid) aux standards du web (WebAnnotation, IIIF) Olivier Aubert analyse dans ce document les formats existants et propose des recommandations pratiques pour l'échange et la préservation des annotations audiovisuelles en sciences humaines et sociales.



Huma-Num IR*

canevas