



MINISTÈRE
DE LA CULTURE

*Liberté
Égalité
Fraternité*

COMITÉ SEDA & CELLULE DE VEILLE : MÉTADONNÉES AUDIOVISUELLES

Plénière ARISTOTE-PIN

La préservation du patrimoine audiovisuel

Mardi 8 octobre 2024 – BnF (Paris)

<https://www.association-aristote.fr/evenements/pin-preservation-du-patrimoine-audiovisuel/>

SEDA

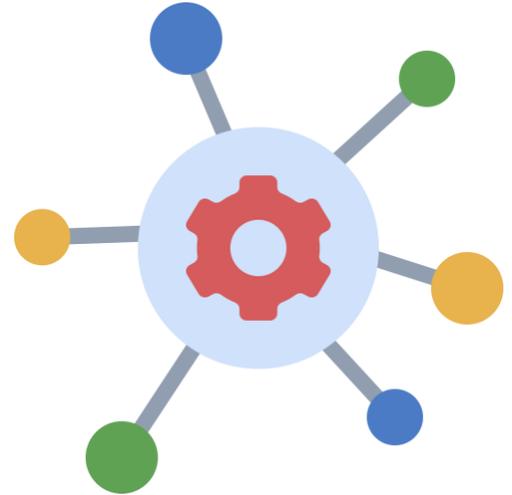
Standard d'échange de données pour l'archivage

SEDA : Favoriser l'interopérabilité entre S.I.

Le standard d'échange de données pour l'archivage vise à faciliter l'interopérabilité entre le système d'information d'un service d'archives public et les systèmes d'informations de ses partenaires.

Sont définis le **format**, la **structure** et le **contenu** informationnel échangés.

<https://github.com/culturecommunication/seda>



La gouvernance du SEDA

Le comité SEDA

- Piloté par le SIAF
- Une diversité d'acteurs (publics et privés) :
 1. éditeurs,
 2. utilisateurs,
 3. consultants,
 4. experts normatifs

Un standard qui évolue au fil des besoins

- v2.2 (Agent générique)
- v2.3 (Signature + Identifiant pérenne)
- v2.4 (2025 : Audiovisuel ?)

Le groupe de travail « *métadonnées audiovisuelles* » du comité SEDA

Objectif

- Améliorer l'archivage des documents audiovisuels numériques (voire numérisés)
- Identifier les métadonnées (descriptives, de gestion et techniques) spécifiques à l'audiovisuel que l'on pourrait incorporer au standard
- Comblent les lacunes de PRONOM (référentiel de formats de fichiers) ?

Fonctionnement

- Créé en mai 2023
- Chantier co-piloté par les Archives nationales et le Programme Vitam
- Une trentaine de membres (services d'archives, ministères, musées, éditeurs, consultants)
- Retex, Parangonnage (AN, BnF, INA, AD34, etc.), extraction de métadonnées techniques

Livrable

- Une spécification détaillée pour conduire une évolution du SEDA

Enrichir SEDA

Exemple des archives audiovisuelles

Identifier le parcours d'utilisation des archives audiovisuelles

Remonter progressivement le courant :

1. Auditer les **pratiques actuelles**
(archives analogiques, numérisées et nativement numérique)
2. Comprendre les **spécificités des documents audiovisuels** (au regard de l'archivage)
3. Etablir une **politique de formats**
4. Cerner les **personae** et leurs besoins respectifs
5. Passer en revue les lacunes des **outils** (fonctionnalités des applications) :
pré-versement / SAE / profils > collecte / conservation / communication
6. Répertorier les **métadonnées dont j'ai besoin** (gestion, description, technique)
7. Isoler les **métadonnées** déjà disponibles
(SEDA... mais aussi METS, PREMIS, DublinCore, MPEG-7, etc.)
8. Définir les métadonnées **à créer et/ou embarquer** dans **SEDA**

Etablir une politique de formats

Aristote – Groupe PIN

Cellule nationale de veille sur les formats (CNVF)

https://www.association-aristote.fr/wp-content/uploads/2022/11/05_Pleniere-PIN-novembre-2022-Actualites-Cellule-Formats.pdf

Retours d'expériences sur les politiques formats

- **Bibliothèque nationale de France (BnF)**

Transmettre et partager les actions de préservation numérique : l'initiative PAR (Preservation Action Registry)

<https://siaf.hypotheses.org/1562>

- **Archives nationales de France (ANF)**

Retour d'expérience sur la stratégie de préservation des Archives nationales

<https://siaf.hypotheses.org/1531>

Une approche centrée utilisateurs

Mes personae :

1. Les usagers
2. Les archivistes
3. Les producteurs d'archives
4. Les applications métiers d'archivage
(ex: pré-versement, SAE, générateur de profil d'archivage, référentiels, etc.)

Décrire les besoins de chaque persona

Rédiger des récits utilisateurs (user stories)

- Quel est le besoin ?
- A quelle étape du cycle de vie de l'archive, le besoin surgit-il (collecte, préservation, accès, etc.) ?
- Quelle est sa fréquence (quotidienne, annuelle, décennale, etc.) ?
- Quelles sont ses criticité et priorité (MoSCoW) ?
- **Quels sont les options envisageables pour le satisfaire :**
Scenario maximal / intermédiaire / minimal ?
- Quelles modalités pour la mise en œuvre de ces scenarii ?

Exemples de cas d'usage

Identifier (et prioriser) les cas d'usages par persona :

- **Collecter des archives**
(ex: contrôle à l'entrée sur les formats grâce à des profils)
- **Interroger les métadonnées**
(instruire une demande d'accès; établir des stats; préparer un audit)
- **Auditer un fonds**
(ex: connaître les formats dans mon SAE et suivre leur évolution sur plusieurs décennies)
- **Exploiter *programmatically* les métadonnées pour transformation en masse des fichiers**
- **Tracer les événements sur un fonds**
(ex: profil de numérisation ou documenter les transcodages)
- **Restauration numérique**
(ex: créer des instances dérivées pour préservation)
- **Instruire une demande d'accès**
(ex: préparer la consultation en salle de lecture)
- **Etc.**

Maîtriser les risques du projet

Est-ce que j'en ai besoin... pour l'archivage ?

- Pourquoi ai-je besoin de connaître cette information ?
- Pourquoi cette information et pas d'autres (pourtant disponibles) ?
- Dans quelle mesure est-elle utile/indispensable aux processus d'archivage ?
- Frugalité / Les métadonnées essentielles

SEDA ou archivage numérique ?

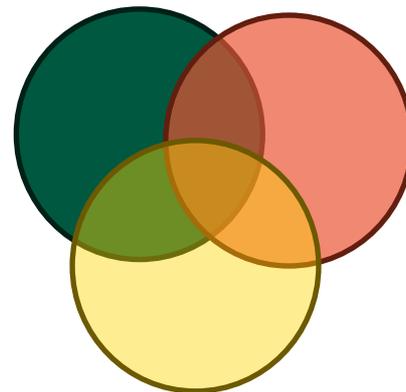
SEDA standardise **les échanges** entre S.I. => interopérabilité !

Toutefois, d'un point de vue **fonctionnel** et **technique**, un S.A.E.* ne repose pas intégralement sur SEDA.

Certains besoins peuvent s'appuyer sur SEDA...
sans pour autant être pris en charge *de bout en bout* par SEDA.

Deux exemples « sans SEDA » :

- **Dans le SAE**
Les Griffons Vitam : FFMPEG, ImageMagick, LibreOffice, etc.
- **Hors du SAE**
Audit de numérisations : programme Python pour interroger les métadonnées techniques avec FFPROBE depuis un PC bureautique et produire un CSV



* SAE : *Système (d'information) d'Archivage Électronique*

Tropisme *archivistique*

- **Attention aux *biais de confirmation* entre pairs !**
- **Créer un groupe de travail pluridisciplinaire :**
 1. Archivistes d'horizons divers (AN, AD, Missions, Musées, etc.)
 2. Producteurs d'archives (ex: Ministères, Collectivités territoriales)
 3. Sociétés de production audiovisuelle
 4. Usagers : Documentariste / Chercheur / Etc.
 5. Editeurs de logiciels (SAE) et développeurs informatiques
 6. DSI

Les personae et cas d'usage sont produits collectivement en prenant en compte les connaissances de chaque activités... pour évaluer l'adéquation de mon scénario à la pratique réelle.

Biais « archives physiques »

- Reproduction « intuitive » de la situation qui prévaut avec l'analogique... alors que le besoin ou le contexte sont potentiellement différents avec les archives numériques ou numérisées.
- Biais issu de la documentation « manuelle » des archives audiovisuelles analogiques. L'archiviste documente lui-même les formats (description).



Pour mémoire : les métadonnées techniques sont embarquées dans le media. Dans une perspective purement archivistique, ai-je besoin de les répliquer sous une autre forme (via extraction et retraitement ?)

Adhérences et aléas

J'ai besoin de **garantir une relative stabilité** de la réponse apportée à mon besoin :

- Quelles *tierces parties* seront sollicitées ?
- Puis-je influencer sur leur gouvernance ?
- Ai-je de la visibilité sur leur feuille de route (évolutions sur les x prochaines années) ?

Par exemple :

- **Référentiels** : codecs audio et vidéo => référentiel(s) existant(s), partagé, public, exhaustif... **ou à créer ?**
- **Applications** : FFMPEG, MediaInfo => licence, dynamisme, format d'export, couverture fonctionnelle ?
- **Normes** : ISO MPEG-7 => version 2 ou version 4 ?

MediaInfo (version 24 – juin 2024) utilise MPEG-7 **v2** (et non **v4**) :

- http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-7_schema_files/mpeg7-v2.xsd
- https://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-7_schema_files/mpeg7-v4.xsd

Commencer petit...

Le contrôle à l'entrée
ou la transformation de format
nécessitent d'abord d'auditer
les codecs.

Quant à la préservation...
le cœur du sujet réside probablement
dans la traçabilité
des **codecs** audio et vidéo.

Développer **un référentiel de
codecs** pour adosser mes
processus métiers sur celui-ci ?

*Référentiel externe ou interne ?
Référentiel complet ou empirique ?*

Merci

Louis Vignaud

Expert chargé de la politique nationale sur les métadonnées et référentiels archivistiques

Bureau de l'expertise numérique et de la conservation durable

Service interministériel des Archives de France

louis.vignaud@culture.gouv.fr