



**HAL**  
open science

## Knowledge about Metadata

Azim Roussanaly, Olfa Messaoud, Asif Ahmed

► **To cite this version:**

Azim Roussanaly, Olfa Messaoud, Asif Ahmed. Knowledge about Metadata. [Technical Report] LORIA - Université de Lorraine. 2021. hal-03625230

**HAL Id: hal-03625230**

**<https://hal.inria.fr/hal-03625230>**

Submitted on 30 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



SmartVideo Grand Est Project

---

# Knowledge about Metadata

---

Azim ROUSSANALY  
Olfa MESSAOUD  
Asif AHMED

Version of  
16 juin 2021

This document is a kind of annex for the SmartVideo Grand Est project. Noting that SmartVideo project consists of conceiving and implementing an on-line streaming platform for Grand Est local channels providing a recommender system to help users of Noozy.tv platform to easily find their interesting and relevant content basing on user experience and previous navigations.

The aim of this document is to describe the metadata about videos or contents provided by local channels producers. In fact, the collaborators opted for adopting the Dublin Core Standard [2] to present knowledge about metadata. Dublin Core was firstly presented in 1995 and is considered as a vocabulary of fifteen properties aiming to present a wide range of resources [3].

The fifteen element Dublin Core described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI). The DCMI presents two broad classes of qualifiers [4] :

- **Element Refinements** : A refined element shares the meaning of the unqualified element, but with a more restricted scope.
- **Encoding Schemes** : Schemes help to identify the interpretation of an element value (type format, standard, etc.).

The approved qualifiers do not constitute a closed set. It is expected that implementers will develop additional qualifiers for use within local applications or specific domains. The table 1 presents the basic Dublin Core Metadata Element Set defined by DCMI :

Element	Element refinement(s)
Date	Alternative
Creator	
Subject	
Description	Table of Contents
	Abstract
Publisher	
Contributor	
Date	Created
	Valid
	Valid
	Available
	Issued
Modified	
Type	
Format	Extent
	Medium
Identifier	
Source	
Language	
Relation	Is Version Of
	Has Version
	Is Replaced By
	Replaces
	Is Required By
	Requires
	Is Part Of
	Has Part
	Is Referenced By
	References
	Is Format Of
	Has Format
Coverage	Spatial
	Temporal
Rights	

**TABLE 1** – Dublin Core Metadata Element Set

## Detailed Description of Videos Metadata

Basing on the Dublin Core Metadata Element Set and focusing on our Studying case, the collaborators have provided a similar set of qualifiers that aim to present the Knowledge about the videos used for SmartVideo project. The SmartVideo Metadata Element have provided additional needed elements beside removing some unneeded elements or refinements.

The table 2 details for more the elements (properties) utilized for describing the videos.

Element	Refinement	Description	Type	Example	Nature
Title		Video Title	Text	A feu doux : Domaine du Haut-jardin #5	Required 1
Creator	Identifiant	Author Identifier	URI	ME0000450	Required *
	Title	Author Name	Text	Maud Bettina-Marie	
Subject	Identifiant	Keyword or Thematic Identifier	URI	TH0007, TAG000510	Required *
	Title	Keyword or Thematic Name	Text	Sport, Environnement, Cuisine, Recette, biathlon	
Description	Abstract	Synopsis or Résumé	Text	Dans les cuisines des restaurateurs locaux, Valentine vous propose un moment de partage avec des personnalités attachantes qui s'expriment dans l'art culinaire	Required 1
Publisher	Identifiant	Channel Identifier	URI	ME0000019	Required *
	Title	Channel Name	Text	Vosges TV	
Contributor	Identifiant	Producer Identifier	URI	ME0000288	Required *
	Title	Producer Name	Text	Moselle TV	

Date	Created	Creation Date	W3C DTF	2021-03-18T09 : 54 :08+01 :00	Required 1
	Available	Publication Date	W3C DTF	2021-04-06T19 : 21 :59+02 :00	Required *
	Modified	Modification Date	W3C DTF	2021-04-06T15 : 51 :15+02 :00	Optional 1
	Issued	Publication Last Date	W3C DTF	2021-04-08T10 : 54 :50+02 :00	Required *
Type	Identifier	Genre Identifier	URI	ME0000007	Required *
	Title	Genre Name	Text	Documentaire	
Format	Extent	Total Duration	ISO 8601	01 :10 : 20	Required 1
Identifier		Video Identifier	Text	kqpIgMPbhY	Required 1
Source	Identifier	Program Identifier	URI	EMI000199	Required 1
	Title	Program Name	Text	Sports week-end	
Relation	is Part of	Link to other video(s)	Text	L2j25bWOFy	Required *
Coverage	Spatial	Video Location	Text	lat : 48.579831, lng : 7.761454	Required 1
Rights	Licence	Licence	Text	Free	Optional 1
Rights Holders	Identifier	Holder Identifier	URI	ME0000026	Optional *
	Title	Holder Name	Text	Vosges TV	
Audience		Video Audience	Text	National	Optional 1

**TABLE 2** – SmartVideo Metadata Element Set

**Remark :**

- 1 : Only one value per property
- \* : Multiple values per property

***Diving Deeper into the Elements Values***

We notice that several fields are defined using an Identifier having an URI format and a Title field containing the important data. So, we are going to explore only the Title field of each element. After a thorough analysis, some refinement values are more defined below :

- **Creator.Title** : 37 Authors, Once per video  
*e.g.* Karim Dridi, Anthony Sap, Lucie Perrin, etc.
- **Subject.Title** : 3528 keyword, Multiple per video  
*e.g.* boulangerie, centre-commercial, champignons, chanteur, indien, industrie, innovation, usine, Culture, Divertissement, Economie, Education / Formation, Environnement, Faits divers, Information, Politique, Santé, Société, Sport, Territoire / Patrimoine, Tradition, etc.
- **Publisher.Title** : 4 Channels, Once per video  
*e.g.* Alsace 20, Canal 32, Moselle TV, Vosges TV.
- **Contributor.Title** : 47 Producers, Once per video  
*e.g.* ANA films, ARGANE PRODUCTIONS, Alpaga Films, Alsace 20, Anoki, Blokk, CV Production, Canal 32, Dora Films, French Connection, Hibou Production, Holden, Hutong productions, Iskra, etc.
- **Type.Title** : 8 Genres, Once per video  
*e.g.* Bande annonce, Captation culturelle, Captation sportive, Documentaire, Fiction, Magazine/Emission, Programme court, Reportage.
- **Source.Title** : 34 Source of the provided videos, Once per video  
*e.g.* A feu doux, Le journal de VosgesTV, La case Doc, Sports week-end, Entre nous, Quon se le dise, Concert / Spectacle vivant, Météo, Bande-annonce, Place publique, etc.
- **RightsHolders.Title** : 40 Rights Holders, Once per video  
*e.g.* Lyon Capitale TV, Machette prouction, Mara Films, Octobre Production, Opéra national de Lorraine, Oxymore, P6 Productions, Pylprod, Real production, TS production, The Kingdom, Tryptique films, etc.

***Introduction to OAI-PMH***

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an application-independent interoperability framework based on meta-



---

data harvesting [1]. OAI-PMH is a service giving a simple technical option for data providers to make their metadata available to services.

There are two classes of participants in the OAI-PMH framework :

- **Data Providers** : administer systems that support the OAI-PMH as a mean of exposing metadata
- **Service Providers** : use metadata harvested via the OAI-PMH as a basis for building value-added services

The OAI-PMH most related keywords are :

- **Harvester** : A harvester is a client application that issues OAI-PMH requests.
- **Repository** : A repository is managed by a data provider to expose metadata to harvesters.
- **Item** : An item is a constituent of a repository from which metadata about a resource can be disseminated.
- **Unique Identifier** : A unique identifier unambiguously identifies an item and is used for extracting metadata from the item.
- **Record** : A record is metadata from a given item expressed in a single format.
- **Set** : A set is an optional construct for grouping items for the purpose of selective harvesting.
- **Selective Harvesting** : Selective harvesting allows harvesters to limit harvest requests to portions of the metadata.

### ***OAI-PMH API Requests***

OAI-PMH requests are expressed as HTTP requests. Requests must be submitted using either the HTTP GET or POST methods. POST has the advantage of imposing no limitations on the length of parameters. Repositories must support both GET and POST methods. There is a single base URL for all requests. In addition to the base URL, all requests consist of a list of keyword parameters, which take the form of key=value pairs.

- **key** : String Verb
- **value** : One of the defined OAI-PMH requests
  - **GetRecord** : Retrieve an individual metadata record for a given item.
  - **Identify** : Retrieve information about a repository.
  - **ListIdentifiers** : Abbreviated form of ListRecords, retrieving only headers than records.

- **ListMetadataFormats** : Retrieve the metadata formats available from a repository.
- **ListRecords** : Harvest records from a repository and return all metadata.
- **ListSets** : Retrieve the set structure, useful for selective harvesting.

The OAI-PMH API developed for SmartVideo Grand Est project contains the corresponding endpoints :

- `http://noozy.fr/oai?verb=ListIdentifiers` : Get all video identifiers (headers)
- `http://noozy.fr/oai?verb=ListIdentifiers&from=2021-01-29` : Get all video identifiers from 29 January 2021
- `http://noozy.fr/oai?verb=ListIdentifiers&from=2021-01-29%2000:00:00` : Get all video identifiers from 29 January 2021 specifying the time zone
- `http://noozy.fr/oai?verb=ListIdentifiers&from=2021-01-29&until=2021-02-29` : Get all identifiers from 29 January to 29 February 2021
- `http://noozy.fr/oai?verb=ListRecords` : Get all metadata about items (with pagination limit = 10)
- `http://noozy.fr/oai?verb=ListRecords&page=1` : Get records of the second page
- `http://noozy.fr/oai?resumptionToken=eyJ2ZXJiIjoiTG1zdFJlY29yZHM iLCJwYWdlIjoxfQ==` : Using resumption token to automatically go to the next page
- `http://noozy.fr/oai?verb=GetRecord&identifier=NrQvNcBkzg` : Return all metadata about a video given its ID
- `http://noozy.fr/oai?set=emi` : Selective harvesting, return metadata by emission or program (pagination limit = 20)
- `http://noozy.fr/oai?set=genre` : Selective harvesting, return metadata by type or genre
- `http://noozy.fr/oai?set=theme` : Selective harvesting, return metadata by subject or thematic
- `http://noozy.fr/oai?set=chaine` : Selective harvesting, return metadata by publisher or channel

The listing 1 shows an example of using the OAI-PMH API to recuperate metadata of a given video (ID = "NrQvNcBkzg").

Listing 1 – Get Record for a given video

```
{
  "results": "1",
  "cursor": 0,
```

```
"request": "verb=GetRecord&&identifiant=NrQvNcBkzg",
5 "responseDate": "2021-06-15T20:56:10+02:00",
  "GetRecord": [
    {
      "header": {
        "identifiant": "NrQvNcBkzg",
10 "datestamp": "2021-01-26"
      },
      "metas": {
        "dc:title": "Nouvelle campagne",
        "dc:identifiant": "NrQvNcBkzg",
15 "dc:subject": [
          {
            "identifiant": "TH0010",
            "title": "Sant "
          },
20 {
            "identifiant": "TAG005880",
            "title": "D pistage Covid-19"
          },
          {
            "identifiant": "TAG004928",
25 "title": "Covid-19"
          },
          {
            "identifiant": "TAG000274",
30 "title": "sant "
          }
        ],
        "dc:description": "La Ville d'Epinal lance
          une nouvelle campagne de d pistage
          massif dans diff rents quartiers de la
          commune pour toucher ceux qui n'oseraient
          pas forc ment se rendre en pharmacie,
          chez leur m decin ou dans un laboratoire
          .",
        "dc:publisher": [
35 {
            "identifiant": "ME0000019",
            "title": "VosgesTV"
          }
        ]
      }
    }
  ]
```

```
    ],  
    "dc:created": "2021-01-26T09:30:55+01:00",  
    "dc:modified": "2021-01-26T09:31:11+01:00",  
    "dc:available": "2021-01-21T00:00:00+01:00",  
    "dc:type": [  
      {  
        "identifier": "ME0000007",  
        "title": "Reportage"  
      }  
    ],  
    "dc:extent": "00:02:30",  
    "dc:source": {  
      "identifier": "EMI000197",  
      "title": "Le journal de VosgesTV"  
    },  
    "dc:coverage": [  
      {  
        "lat": 48.17006,  
        "lng": 6.484474  
      }  
    ],  
    "dc:licence": "free",  
    "dc:rightsHolder": [  
      {  
        "identifier": "ME0000026",  
        "title": "VosgesTV"  
      }  
    ],  
    "dc:audience": "departement"  
  }  
}  
]  
}
```

## Conclusion

The SmartVideo Metadata Element Set is elaborated basing on the Dublin Core Standard that is considered as the best standard to describe thoroughly all the needed fields aiming to present completely a video.

## References

- [1] Carl LAGOZE et Herbert Van de SOMPEL : The open archives initiative protocol for metadata harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>, 2002.
- [2] Stuart WEIBEL, John KUNZE, Carl LAGOZE et Misha WOLF : Dublin core metadata for resource discovery. *Internet Engineering Task Force RFC*, 2413(222):132, 1998.
- [3] Stuart L WEIBEL et Traugott KOCH : The dublin core metadata initiative. *D-lib magazine*, 6(12):1082–9873, 2000.
- [4] Stuart L. WEIBEL et Traugott KOCH : The dublin core metadata initiative. <http://mirror.dlib.org/dlib/december00/weibel/12weibel.html#ref3>, 2000.



LORIA  
Campus Scientifique  
BP 239  
54506 Vandoeuvre-lès-Nancy