

LES DÉFIS DE LA NUMÉRISATION DES ARCHIVES AUDIOVISUELLES DE CANAL9 (VALAIS, SUISSE) ENTRE COLLECTE, PRÉSERVATION ET VALORISATION

Alain Dubois

RESUMEN

Situado en el corazón de los Alpes suizos, el cantón de Valais acoge en su territorio el canal de televisión privado más antiguo de Suiza. Fundado en 1984 en Sierre, el Canal9 produjo emisiones en analógico hasta 2005, momento en que se pasó al medio digital. El depósito de archivos audiovisuales de la Mediateca de Valais fue en 2015 el punto de partida de un proyecto piloto para la preservación de un recurso esencial de historia regional financiado por la Oficina Federal de Comunicación y Memoria. Se han tenido que afrontar varios retos, y otros lo serán próximamente: elaboración de un modelo de paquete de información compatible con el modelo de referencia OAIS (Matthhorn METS Profile); desarrollo de una herramienta de reconocimiento de la voz, capaz de gestionar el bilingüismo del cantón (franco alemán); alineamiento del corpus en estándares de datos Web, de modo que se pueda enriquecer y enlazar con conjuntos más grandes o incluso configurar un proceso de archivo más cercano a la producción de emisiones por parte del corpus creado en digital.

SUMMARY

Located in the heart of the Swiss Alps, the canton of Valais broadcasts, throughout its area, the oldest private television channel in Switzerland. Founded in 1984 in Sierre, Canal9 produced analogue programs until 2005 when it transitioned to digital media. In 2015, a pilot project to preserve this essential source of regional history was launched by the Valais Mediatheque with the support of the Federal Office of Communication and Memoriav. Several challenges have been or will soon be overcome: development of an information package model which is compliant with the OAIS reference model (Matthhorn METS Profile); development of a speech recognition tool; the ability to work with the canton's bilingualism (French-German); confirm that the corpus is in line with Web data standards in order to be expanded and linked to larger sets; and establish an archiving process which will digitally archive broadcasts upon their production, or as soon as possible afterwards.

RESUM

Situat al cor dels Alps suïssos, el cantó de Valais acull al seu territori el canal de televisió privat més antic de Suïssa. Fundat el 1984 a Sierre, el Canal9 va produir emissions en analògic fins al 2005, moment en què es va passar al mitjà digital. El dipòsit del arxius audiovisuals de la Mediateca de Valais van ser el 2015 el punt de partida d'un projecte pilot per a la preservació d'un recurs essencial d'història regional finançat per l'Oficina Federal de Comunicació i Memòria. S'han hagut d'afrontar diversos reptes, i altres ho seran properament: elaboració d'un model de paquet d'informació compatible amb el model de referència OAIS (Matthhorn METS Profile); desenvolupament d'una eina de reconeixement de la veu, capaç de gestionar el bilingüisme del cantó (francoalemany); alineament del corpus en estàndards de dades web, de manera que es pugui enriquir i enllaçar amb conjunts més grans o fins i tot configurar un procés d'arxiu més proper a la producció d'emissions per part del corpus creat en digital.

Situé au cœur des Alpes suisses, le canton du Valais accueille sur son territoire la plus ancienne chaîne de télévision régionale de Suisse. Fondée en 1984 à Sierre, Canal9 a produit des émissions sur support analogique jusqu'en 2005 avant de migrer vers le support électronique à partir de cette date. Le dépôt des archives audiovisuelles à la Médiathèque Valais en 2015 a été le point de départ d'un projet pilote de préservation d'une source essentielle pour l'histoire régionale soutenu notamment par l'Office fédéral de la communication et Memoriav. Plusieurs défis ont dû être relevés ou le seront dans les prochains mois : élaboration d'un modèle de paquet d'information conforme au modèle de référence OAIS (Matthhorn METS Profile), développement d'un outil de reconnaissance de la parole, capable de gérer le bilinguisme du canton (français-allemand), alignement du corpus sur les standards du Web de données, de manière à pouvoir l'enrichir et le relier à des ensembles plus vastes, ou encore mise en place d'un processus d'archivage au plus près de la production des émissions pour la partie du corpus née numérique. C'est ainsi de la réalisation de ce projet que traite le présent article.

La télévision régionale Canal9 en bref

Imaginée en 1973 par Marc Biderbost, animateur au sein de l'Association sieroise de loisirs et de culture (ASLEC), la chaîne de télévision Canal9 a été concrètement fondée en 1984 par la Jeune chambre économique de Sierre sous le nom de Sierre Canal9. Elle est la plus ancienne chaîne de télévision régionale de Suisse.

D'abord diffusée dans la région de Sierre, située dans le Valais central, Canal9 étend en 2002 son périmètre de diffusion à toute la partie francophone du Valais. En septembre 2005, Canal9 devient la première télévision régionale européenne à utiliser uniquement des moyens informatiques pour assurer toute sa production. C'est également à cette date que les studios sont regroupés à Sierre, tandis que des antennes régionales sont créées à Monthey, à Martigny, à Sion et à Brigue et sont reliées entre elles par fibre optique.

Le 31 octobre 2008, Canal9 obtient de l'Office fédéral de la communication la concession pour diffuser ses programmes en Valais face au projet concurrent Valais-Wallis TV. Le 29 juin 2009, elle diffuse sa première émission en langue allemande et assure désormais une diffusion sur l'ensemble du territoire du canton du Valais. Elle dispose à partir de ce moment-là d'une rédaction de langue française située à Sierre et d'une rédaction de langue allemande hébergée à Brigue.

Canal9 diffuse depuis le 9 septembre 2013 la totalité de sa programmation en haute définition (HD). Il est également possible de visionner ses émissions en direct ou en différé sur son site internet et les applications mobiles¹.

Média de service public, Canal9 réalise ainsi chaque soir de la semaine des émissions d'actualité et des magazines qui reflètent à la fois la vie du canton, dans ses dimensions politiques, économiques, culturelles ou sociales, et les préoccupations de ses habitants².

Un patrimoine en danger

Élément important, Canal9 a conservé de manière systématique ses archives depuis le début de ses activités en 1984, en confiant la gestion de ces dernières à une équipe d'archivistes professionnels. L'ensemble constitue un patrimoine audiovisuel de première importance pour l'histoire du canton du Valais des 35 dernières années, qui représente un complément essentiel et indispensable aux sources administratives de l'Etat du Valais et des communes valaisannes, ainsi qu'aux fonds d'archives des personnes privées, physiques ou morales, disponibles pour cette période³. De 1984 à 2005, les émissions de Canal9 ont été en partie conservées sur cassettes U-Matic et surtout sur cassettes Betacam⁴. Depuis septembre 2005, le système de production est informatisé ; le format d'archivage retenu est le format AVI DV qui est aujourd'hui réputé pérenne.

Par suite de l'arrêt du financement des télé réseaux du Valais romand, Canal9 a connu en 2013-2014 d'importantes difficultés financières qui l'ont contraint à introduire des mesures de restructuration et d'économies, dont la fermeture du secteur des archives et le licenciement de l'équipe des archivistes. Malgré tout, Canal9 a pris rapidement des mesures pour assurer la préservation et la transmission aux générations à venir de ses archives audiovisuelles et a décidé en 2014 de les déposer à la Médiathèque Valais-Martigny, institution culturelle en charge notamment de la préservation de la mémoire audiovisuelle valaisanne. Une convention de dépôt a ainsi été signée dans la

1. [En ligne :] <https://www.canal9.ch> (consulté le 22 septembre 2018).

2. Pour des compléments sur l'historique de Canal9, se référer à [en ligne :] <http://www.mediathèque.ch/valais/archives-canal9-5125.html> (consulté le 22 septembre 2018)

3. S'il est du reste fréquemment utilisé par des vidéastes ou des journalistes pour illustrer leurs propos, le fonds n'a pas encore fait l'objet d'une exploitation systématique par les chercheurs. Sa numérisation conduira certainement à une exploitation systématique par le monde académique, dans le sillage des projets similaires lancés en Suisse sur ce type de corpus.

4. La répartition des cassettes est la suivante : 1630 cassettes U-Matic et 5060 cassettes Betacam.

foulée entre les deux parties pour garantir une prise en charge du corpus conforme aux normes et standards de la conservation du patrimoine.

Le plus grand risque auquel étaient confrontées les archives audiovisuelles de Canal9 en 2014 résidait avant tout dans l'obsolescence des supports de conservation de l'information. Aucune détérioration physique majeure des cassettes, telle que moisissures ou rétrécissement, n'a par contre été constaté au moment de l'analyse du corpus, les conditions de conservation ayant été globalement satisfaisantes. L'état sanitaire du fonds était ainsi bon.

Un projet pilote national pour préserver un patrimoine unique

La signature de la convention de dépôt des archives de Canal9 à la Médiathèque Valais constitue le socle sur lequel s'est construit un projet pilote d'envergure nationale pour préserver ce patrimoine unique.

Le groupe de projet interne constitué par la Médiathèque Valais a effectué une analyse du corpus sous l'angle archivistique. Il a tout d'abord décidé de limiter le projet, pour des questions stratégiques et pratiques, dans une première phase (2015-2017), aux archives conservées sur support analogique. Il a de même été décidé d'appliquer des critères d'évaluation stricts au regard des 6700 cassettes conservées pour la période comprise entre 1984 et 1985 et de ne finalement traiter et préserver que les 3000 émissions pour 3300 heures produites au cours de cette période, qui constituent la forme la plus aboutie de la documentation disponible. Les rushes ou images brutes, pour leur part, qui peuvent avoir un intérêt certain, notamment les phénomènes climatiques extrêmes, tels que des inondations ou des crues avalanches⁵, ou les événements majeurs pour l'histoire du canton du Valais, ainsi que les captations de manifestations sportives, culturelles ou politiques⁶ feront l'objet d'une évaluation et d'un traitement spécifiques dans une nouvelle phase du projet. Les questions de description et de mise à disposition du corpus ont fait l'objet de nombreuses réflexions et ont débouché sur la mise en place de solutions innovantes présentées ci-dessous. Pour le surplus, les notices descriptives sont intégrées au catalogue de la Médiathèque Valais.

5. Le canton du Valais a connu plusieurs phénomènes climatiques extrêmes au cours des dernières décennies : inondations à Brigue en 1993 et dans la plaine de Rhône en 1999 ou encore crue avalancheuse sur l'ensemble du territoire en 2000.

6. Parmi les événements particuliers, il convient de citer les championnats du monde de ski en 1987 à Crans-Montana ou la visite du Conseil fédéral en 2003.

La spécificité du projet de Canal9 réside dans le vaste réseau mis en œuvre pour assurer la préservation de ce patrimoine. De nombreux partenaires ont ainsi été associés au projet. Ce dernier a pu tout d'abord bénéficier de l'engagement et de l'expérience de Memoriav, l'association pour la sauvegarde de la mémoire audiovisuelle suisse, qui a déjà soutenu plusieurs projets de préservation de fonds d'archives audiovisuelles de diffuseurs privés de télévision⁷. Outre des moyens financiers, Memoriav a ainsi pu amener un réseau de compétences et des recommandations bienvenues dans le cadre du projet. Représentée par l'Office fédéral de la communication (OFCOM), la Confédération helvétique a par ailleurs apporté un soutien essentiel à un projet qui s'inscrit dans les objectifs qu'elle entend actuellement poursuivre. Le message culture 2016-2020 du Conseil fédéral, d'une part, considère les archives des diffuseurs suisses de radio et de télévision comme un pilier essentiel du patrimoine audiovisuel suisse et propose un certain nombre d'actions pour le préserver et le mettre à la disposition de la population⁸. Le législateur, d'autre part, a récemment reconnu l'importance de la préservation des archives audiovisuelles dans le cadre de la révision de la loi fédérale sur la radio et la télévision du 24 mars 2006 (LRTV⁹) et dans l'ordonnance du 2 mars 2007 qui s'y rattache (ORTV¹⁰). L'art. 33, al. 1, de l'ordonnance sur la radio et la télévision énonce notamment que l'Office fédéral de la communication « peut soutenir des projets de conservation durable des émissions » produites par des diffuseurs privés notamment titulaires d'une concession. C'est ainsi que Canal9 a été la première chaîne de télévision privée à bénéficier de ce nouveau dispositif législatif et à recevoir d'importants moyens financiers pour préserver ses archives audiovisuelles. La contrepartie de ce soutien réside dans la mise à disposition des archives : « Les émissions qui ont été conservées durablement avec le soutien de l'OFCOM doivent être rendues accessibles au public sous une forme appropriée pour un usage privé ou scientifique, dans le respect des droits de tiers (art. 33, al. 2) ». C'est ainsi que, conformément aux souhaits de l'Office fédéral de la communication, le groupe de projet a développé

7. [En ligne :] <https://www.memoriav.ch> (consulté le 22 septembre 2018).

8. Message concernant l'encouragement de la culture pour la période 2016-2010 du 28 novembre 2014. [En ligne :] <https://www.admin.ch/opc/fr/federal-gazette/2015/461.pdf>, p. 535-540 (consulté le 22 septembre 2018).

9. [En ligne :] <https://www.admin.ch/opc/fr/classified-compilation/20001794/index.html> (consulté le 22 septembre 2018).

10. [En ligne :] <https://www.admin.ch/opc/fr/classified-compilation/20063007/index.html> (consulté le 22 septembre 2018).



Page d'accueil du site Internet consacré aux archives audiovisuelles de Canal 9.

un portail accessible au public dont les spécificités seront présentées ci-dessous¹¹.

Afin de respecter les droits de tiers, le contenu est du reste visible uniquement en streaming ; le téléchargement, pour sa part, est possible, mais il est soumis à l'autorisation préalable de Canal 9 et au versement d'une contribution financière au titre du droit d'auteur.

Ont également été associés au projet les Archives de l'Etat du Valais et un prestataire de service, en charge de la définition des processus de numérisation et de la mise en œuvre d'un processus d'archivage sur la plateforme de pérennisation et d'archivage de la Médiathèque Valais. Enfin, dans la mesure où il constitue un projet pilote au niveau suisse, le projet a fait l'objet d'un suivi qualitatif assuré par la filière information documentaire de la Hochschule für Wirtschaft und Technik de Coire.

Au final, le groupe de projet a établi le programme de travail prévisionnel suivant :

- numérisation du corpus : 2016-2017
- indexation et catalogage du corpus : 2017
- mise en valeur du corpus à travers un portail dédié : 2017

Ambitieux, ce programme prévisionnel a pu être respecté. Le projet fait actuellement l'objet d'une seconde phase en lien avec le traitement des archives disponibles sous forme numérique depuis 2005 et la mise en place d'un processus d'archivage automatisé pour la production courante.

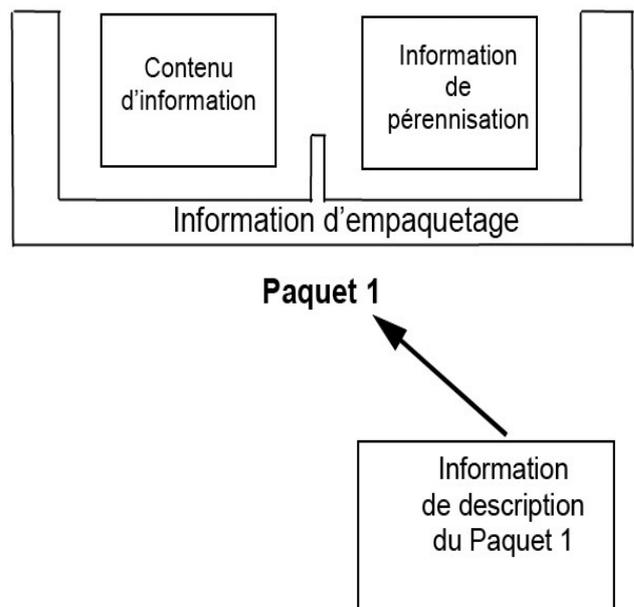
Associé aux autres projets de préservation numé-

rique du Service de la culture, dont font partie les Archives de l'Etat du Valais et la Médiathèque Valais, le projet de traitement des archives audiovisuelles de Canal 9 a permis de mettre en place un certain nombre de processus et d'outils innovants qu'il convient désormais de présenter.

Un modèle de données pour la préservation numérique

La numérisation des archives audiovisuelles de Canal 9 a pour objectif premier de maintenir la lisibilité d'un corpus important sur l'histoire valaisanne de la fin du XX^e siècle et du début du XXI^e siècle pour les générations à venir. Si la numérisation permet de pallier l'obsolescence des supports analogiques utilisés jusqu'en 2005, la conservation du produit issu de la numérisation pose un défi majeur.

Le Service de la culture du Canton du Valais a inauguré en 2011 une plateforme de pérennisation et d'archivage conçue conformément au modèle de référence OAIS et fondée sur des outils open source qui permettent de maîtriser l'ensemble du cycle de vie du processus d'archivage de l'Ingest à la diffusion des archives auprès des publics les plus larges possibles. Cette plateforme accueille aujourd'hui non seulement des fonds d'archives électroniques issues des différentes unités administratives de l'Etat du Valais, mais également les archives audiovisuelles conservées par la Médiathèque Valais-Martigny, dont les archives de Canal 9. L'enjeu principal s'est toutefois situé dans le



Le modèle de paquet d'information tel que défini par le modèle de référence OAIS.

11. [En ligne :] <http://www.mediathèque.ch/valais/archives-canal9-5125.html> (consulté le 22 septembre 2018).

choix d'un modèle d'information pérenne et conforme aux normes et standards internationaux. Selon le modèle de référence OAIS, il est nécessaire non seulement de conserver le contenu d'information en tant que tel, mais également l'information de description, qui permet de connaître le contenu de cette information, et l'information de pérennisation, qui permet de conserver cette information sur le long terme.

A cet effet, les Archives de l'Etat du Valais ont collaboré avec la société Docuteam GmbH pour développer un modèle de données qui puisse répondre aux objectifs de préservation numérique autant pour les Archives de l'Etat du Valais en tant que telles, que pour la Médiathèque Valais. Ce modèle de données, qui se fonde sur des normes et standards ouverts et largement diffusés auprès de la communauté professionnelle, a été enregistré en novembre 2012 auprès de la Bibliothèque du Congrès, à Washington, sous le nom de Matterhorn METS Profile¹². Il utilise ainsi différents standards pour la description d'un paquet d'information constitué dans le cadre de l'archivage de Canal9 : Dublin Core pour la description de l'information de description¹³, PREMIS pour la description de l'information de pérennisation¹⁴ et METS en tant que format conteneur¹⁵. Pour des raisons pratiques, le groupe de travail a décidé que chaque émission de Canal9 constituerait un seul et même paquet d'information ; il lui paraissait par contre peu pertinent d'abaisser cette granularité au niveau d'un sujet traité au cours d'une émission.

Les archives numérisées de Canal9 sont actuellement toutes déposées sur la plateforme de pérennisation et d'archivage de la Médiathèque Valais, encodées conformément au Matterhorn METS Profile. Elles représentent près de 30 TB de données.

Vers la définition d'un modèle de données aligné sur les standards du web de données

Les archives audiovisuelles de Canal9 sont aujourd'hui accessibles à travers un moteur de recherche classique mis à la disposition du public par la Médiathèque Valais-Martigny et dont la page d'accueil a été présentée ci-dessus¹⁶.

Cette mise en valeur de premier niveau s'accompagne néanmoins depuis plusieurs mois d'une réflexion approfondie sur l'alignement du corpus aux standards du web des données. L'objectif est de permettre d'intégrer les archives audiovisuelles de Canal9 à des ensembles plus vastes et de mieux les relier à des fins de sérendipité. Le groupe de travail développe ainsi depuis plus d'une année un nouveau modèle de données appelé Matterhorn RDF Data Model qui propose une description native des archives électroniques en RDF et constitue un développement complémentaire du Matterhorn METS Profile¹⁷. L'utilisation de cette technologie doit non seulement permettre d'intégrer les archives de Canal9 à des référentiels largement diffusés au niveau international, à l'exemple de Wikidata, mais également de les relier à d'autres corpus de sources de l'histoire valaisanne, de manière à créer des corpus entièrement intégrés entre eux. Cent cinquante années de presse écrite en Valais sont depuis peu, par exemple, entièrement accessibles en ligne¹⁸. Pouvoir relier ces deux corpus non seulement du point de vue de leur description, mais surtout de leur contenu constitue un défi majeur qui ne peut être réalisé qu'en utilisant les standards du web de données. Offrir la possibilité au public de rechercher les sources en lien avec une thématique spécifique dans la presse écrite et la presse audiovisuelle constitue néanmoins un défi stimulant sur lequel le groupe travaille depuis près d'une année et qui devrait apporter des résultats d'ici quelques mois.

Vers un archivage en temps réel ?

Un des objectifs du projet est de mettre en place un processus automatisé d'archivage des émissions de Canal9 au plus près de la production, de manière à garantir la constitution d'un corpus de qualité. Actuel-

12. [En ligne :] <https://www.loc.gov/standards/mets/news112912.html> (consulté le 22 septembre 2018). Sur le Matterhorn METS Profile, se référer en particulier à Alain Dubois, « Une plateforme pour la pérennisation du patrimoine valaisan », dans Delphine Debons (éd.), *Des institutions au service du patrimoine culturel. Changement et continuité dans le contexte valaisan / Institutionen im Dienst des Kulturerbes. Wandel und Kontinuität im Walliser Kontext*, Sion, 2018 (Cahiers de Vallesia 29), p. 290-300

13. Le choix du Dublin Core et non de l'EAD, utilisé pour la description archivistique, s'est justifié par le standard de description utilisé par la Médiathèque Valais-Martigny, qui se fonde sur les normes et standards de la bibliothéconomie. Se référer à [en ligne :] <http://dublincore.org/> (consulté le 22 septembre 2018).

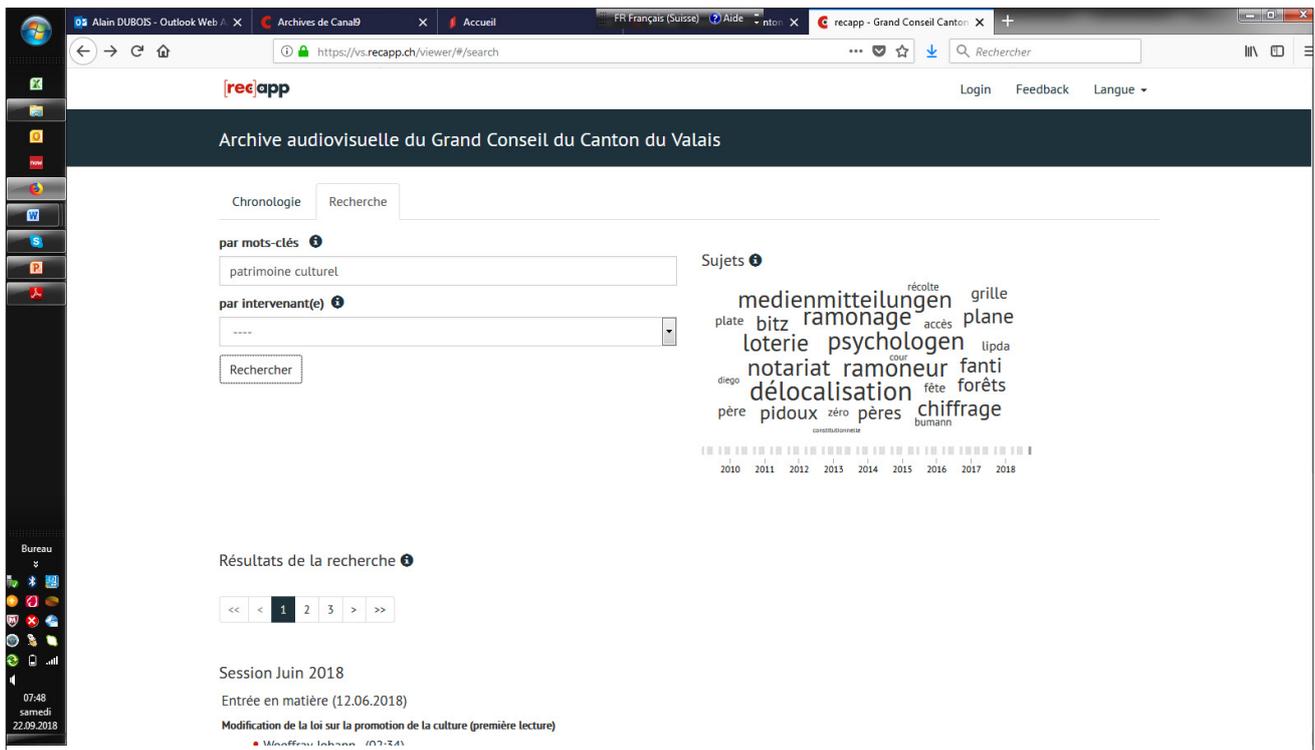
14. Data Dictionary for Preservation Metadata. Se référer à [en ligne :] <https://www.loc.gov/standards/premis/> (consulté le 22 septembre 2018).

15. Sur METS, se référer à [en ligne :] <http://www.loc.gov/standards/mets/> (consulté le 22 septembre 2018).

16. [En ligne :] <http://www.mediathèque.ch/valais/archives-canal9-5125.html> (consulté le 22 septembre 2018).

17. Une première présentation du Matterhorn RDF Data Model a été faite en 2017 dans le cadre du Forum des archivistes genevois et lors de la réunion annuelle du Conseil international des archives à Mexico. Se référer à [en ligne :] <https://www.slideshare.net/forumdesarchivistes/le-matterhorn-rdf-data-model-80360789> et <http://www.alaarchivos.org/wp-content/uploads/2017/12/3.-Alain-Dubois-Andreas-Nef.pdf> (consulté le 22 septembre 2018).

18. Se référer à [en ligne :] <https://www.e-newspaperarchives.ch/?l=fr> (consulté le 22 septembre 2018).



Moteur de recherche consacré aux archives audiovisuelles du Grand Conseil valaisan.

lement en cours de discussion, ce processus vise à permettre à Canal9 de déposer chaque jour le contenu de l'émission de la veille sur la plateforme de pérennisation et d'archivage. Cet objectif ambitieux nécessite néanmoins de résoudre les défis suivants :

- assurer une conservation des archives audiovisuelles à un coût raisonnable. Celles-ci seront ainsi conservées sur des bandes LTO et non plus sur des serveurs ;
- sérialiser les paquets d'information au sens du modèle de référence OAIS. Il s'agit de pouvoir automatiquement encapsuler dans un paquet d'information les données et les métadonnées provenant des outils de production de Canal9 ;
- créer automatiquement des paquets d'information sur la base des outils de la plateforme de pérennisation et d'archivage. Cet élément nécessite le redéploiement de certains outils pour assurer des temps de traitement raisonnables malgré la volumétrie des données.

Les processus mis en œuvre au cours de cette seconde étape font actuellement l'objet d'une analyse approfondie. Quoiqu'il en soit, l'intégration de 30 TB d'archives audiovisuelles supplémentaires chaque année selon les normes et standards de la préservation numérique est loin d'être anodine, notamment en termes de temps de traitement des données. L'infrastructure actuellement mise en place sur la plate-

forme de pérennisation et d'archivage doit ainsi être redéfinie et redéployée pour répondre à cet objectif de pérennisation.

La reconnaissance de la parole

Ne disposant pas des ressources financières et temporelles suffisantes pour assurer une indexation manuelle du corpus, le groupe de travail a rapidement pris la décision d'utiliser les technologies de l'information et de la communication pour l'accompagner dans ces démarches. Il a ainsi collaboré avec recapp, une entreprise issue de l'Idiap Research Institute de Martigny et spécialisée dans la reconnaissance de la parole et la retranscription des données grâce à l'utilisation du *machine learning*¹⁹.

Développée dans le cadre d'une thèse de doctorat, l'algorithme est utilisé depuis plusieurs années pour retranscrire les débats du Grand Conseil du Canton du Valais²⁰. Les autorités législatives ont en effet souhaité pouvoir disposer d'une retranscription complète des débats à la fin de la session. L'application qui en résulte est capable de reconnaître et d'indexer les

19. Se référer à [en ligne :] <https://www.recapp.ch/fr/> (consulté le 22 septembre 2018).

20. Se référer à [en ligne :] <https://vs.recapp.ch/viewer/#/> (consulté le 22 septembre 2018).

interventions faites par les députés dans les deux langues officielles du canton du Valais, le français et l'allemand selon le principe suivant : retranscription des interventions de chaque député, indexation dans une banque de données de textes et identification précise de chaque mot. Surtout, l'application permet désormais aux secrétaires de gagner un temps considérable, puisque l'opération de transcription, chronophage, est supprimée au profit d'une simple relecture. Il s'agit, au surplus, d'un outil qui favorise la transparence de l'action des autorités politiques.

La personne intéressée par une thématique particulière insère ses mots-clés – en français et/ou en allemand – dans le moteur de recherche développé pour l'occasion ; une liste de résultats s'affiche.

L'utilisateur a ensuite la possibilité de visionner la séquence de son choix. Chaque fois que les mots-clés sont mentionnés, une étoile apparaît dans la frise chronologique qui déroule la séquence, pour une meilleure visualisation des résultats de recherche.

C'est ainsi cette même philosophie qui a été retenue dans le cadre des archives audiovisuelles de Canal9.

L'entreprise mandatée a ainsi divisé ce projet de reconnaissance de la parole en quatre phases réalisées entre 2016 et 2017 :

– phase de *Proof of Concept*. Cette première phase avait pour objectif de déterminer les forces et les fai-

bles de l'algorithme développé par recapp par rapport à un échantillon d'émissions de Canal9. Elle avait également pour objectif de définir précisément le processus de traitement automatisé de la parole : mise à disposition par la Médiathèque Valais de quelques émissions, y compris les informations en lien avec le séquençage des contenus de chaque émission et les différents intervenants (métadonnées remises sous la forme d'un fichier XML), importation des données et des métadonnées dans le logiciel développé par l'entreprise mandatée, y compris extraction du signal audio et conversion des métadonnées dans un format propre, traitement des données et des métadonnées, à travers la transcription du signal audio, et exportation des résultats par la remise d'un fichier XML. Cette première phase a nécessité quelques jours de travail ;

– Phase pilote. Cette phase avait pour objectif principal d'automatiser le processus mis en place dans le cadre de la phase précédente et de remédier aux éventuels problèmes rencontrés. L'entreprise mandatée a par ailleurs développé une interface utilisateur pour les collaborateurs de la Médiathèque Valais-Martigny chargés de transcrire 20 heures d'émissions afin de les intégrer à l'algorithme et d'améliorer les performances de ce dernier. Cette seconde phase a nécessité près d'un mois de travail.

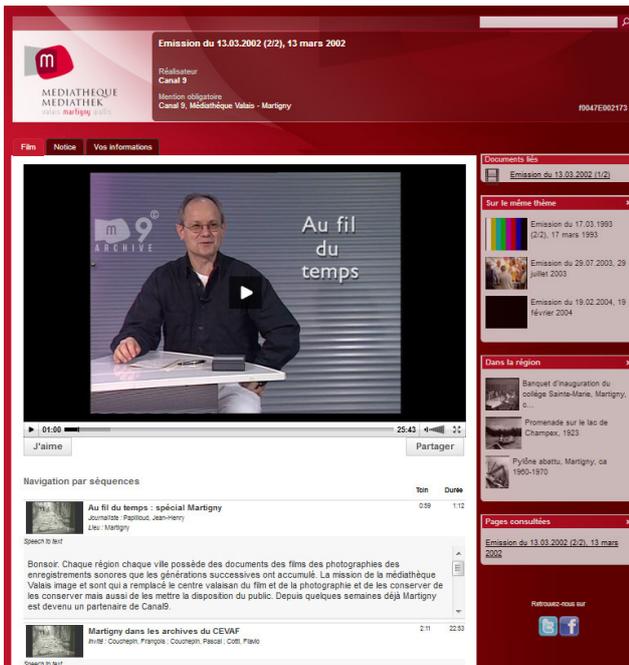
– Phase d'adaptation. Fondée sur les résultats de la phase précédente, cette phase avait pour objectif

The screenshot shows a web browser window displaying the recapp website. The page title is "Entrée en matière" and the subtitle is "Modification de la loi sur la promotion de la culture (première lecture)". The main content area features a video player with a thumbnail of a legislative session. To the right of the video, there is a list of interventions with the following details:

- Dubuis Alexandre
- 12.06.2018, Durée: 2:16
- Interventions
- Schwestermann Alex (1:19)
- Michalet Audrey (2:24)
- Zenhäusern Marcel (2:12)
- Dubuis Alexandre (2:16)**
- Boand Vincent (1:42)
- Birbaum Thomas (0:55)
- Constantin Patricia (2:08)

Below the list is a "liste complète" button. At the bottom of the page, there is a "Mots-clés" section with a horizontal timeline from 0:00 to 2:10. The keywords "patrimoine" and "culturel" are highlighted with blue bars and asterisks along the timeline, indicating their occurrence in the video.

Page en lien avec le résultat d'une recherche sur le patrimoine culturel.



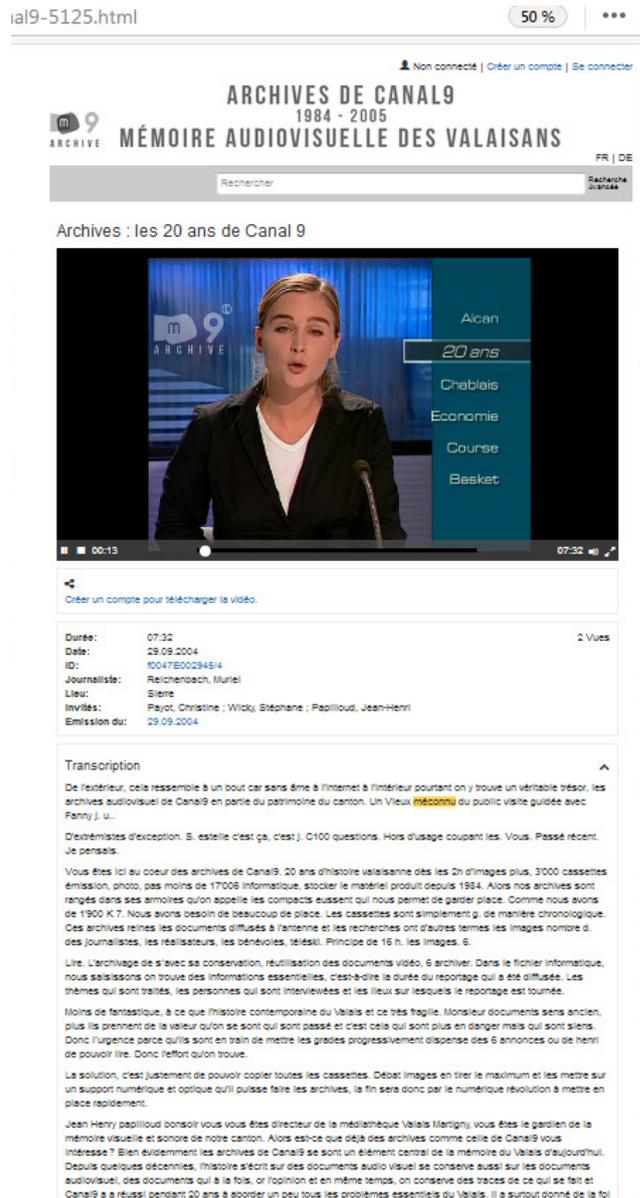
Emission de Canal9 consacrée au patrimoine audiovisuel valaisan (13 mars 2002).

d'adapter le système recapp aux spécificités des archives audiovisuelles de Canal9 et d'améliorer ainsi de manière significative le taux de reconnaissance de la parole. C'est ainsi que le modèle linguistique a été revu et corrigé sur la base des 20 heures d'émissions retranscrites par les collaborateurs de la Médiathèque Valais. De même, le modèle acoustique a été amélioré, en prenant notamment en considération les micros utilisés par les journalistes de Canal9 ou la qualité d'enregistrement. L'ajout de 100 heures d'émissions supplémentaires a contribué à la consolidation des modèles linguistique et acoustique. Cette troisième phase a nécessité près de deux mois de travail.

– Phase de production. Une fois l'algorithme calibré, les 3300 heures d'émissions de Canal9 ont pu être retranscrites et indexées pour une mise à disposition du public.

Le flux XML transmis par recapp une fois le traitement réalisé est, d'une part, intégré au moteur de recherche dédié au patrimoine audiovisuel valaisan et développé par la Médiathèque Valais-Martigny. La personne qui consulte l'émission de son choix peut ainsi accéder à la fois au contenu même de l'émission et à la retranscription de cette dernière, de manière à pouvoir effectuer une recherche plein texte. En voici un exemple avec l'émission du 13 mars 2002 consacrée au patrimoine audiovisuel conservé par la Médiathèque Valais-Martigny.

Le flux XML est intégré, d'autre part, au moteur de recherche développé conjointement par Canal9, la



Emission de Canal9 consacrée à son vingtième anniversaire (20 septembre 2004).

Médiathèque Valais et recapp. Lorsque la séquence est déroulée, les mots prononcés par le journaliste sont surligné en jaune pour un meilleur repérage.

Au final, la première phase du projet de préservation des archives audiovisuelles de Canal9 a permis de traiter les archives conservées sur support analogique entre 1984 et 2005. Au-delà de l'aspect classique d'un projet de numérisation, ce projet est intéressant pour les éléments innovants auxquels il a fait appel dans le domaine de la préservation numérique, avec un modèle de paquet d'information conforme aux normes et standards internationaux, dans le domaine de l'indexation du corpus, grâce à un algorithme de reconnaissance de la parole, dans le domaine de la mise en valeur, à travers la définition d'un modèle de

données adapté aux normes et standards du web de données, pour une meilleure intégration du corpus aux référentiels existants.

Le projet de traitement des archives audiovisuelles de Canal9 connaîtra de nouveaux développements à travers une seconde phase qui prendra place à partir de 2019. Celle-ci traitera non seulement les archives numériques à partir de 2005, mais mettra également en œuvre un processus d'archivage en continu.

Au final, les archives audiovisuelles de Canal9 représentent un potentiel d'exploitation important de la part des chercheurs en sciences humaines et sociales, qui bénéficient ainsi d'un corpus complémentaire aux sources d'archives classiquement mises en œuvre dans le cadre de projets de recherche. Surtout, ces archives audiovisuelles doivent permettre à la population valaisanne de mieux s'approprier ou de mieux se réapproprier sa propre histoire et celle de la région dans laquelle elle vit au cours des 35 dernières années. Gageons que ce sera rapidement le cas.

El text de les experiències es pot consultar a:
El texto de las experiencias se puede consultar en:
The texts containing the experiences can be consulted to:
http://www.girona.cat/sgdap/cat/jornades_actes.php