



CANTON DU VALAIS
KANTON WALLIS

Département de la santé, des affaires sociales et de la culture
Service de la culture
Archives de l'Etat du Valais

Departement für Gesundheit, Soziales und Kultur
Dienststelle für Kultur
Staatsarchiv Wallis



Rapport

Destinataire Projektgruppe ViaCar/CARI
Auteur Florian Vionnet, archiviste-records manager, Archives de l'Etat du Valais
Date 25.01.2016 (version 2.0)

KOST Projekt ViaCar/CARI

Paketierung und Ablieferungsformat

1. Introduction

1.1. Buts du document

L'objectif de ce document est de faire un état des lieux ainsi que des propositions au groupe de travail concernant les formats d'archivage et de versement pour les données issues des systèmes ViaCar et CARI. Il vise en particulier :

1. À définir le format d'archivage pour les données issues des systèmes ViaCar et CARI.
2. A définir le ou les formats pour le paquet d'information à verser (SIP) dans les systèmes d'archivage électroniques.

Ces propositions sont encore ouvertes et pourront faire l'objet d'une discussion lors de prochaines séances ainsi qu'au niveau de la KOST.

1.2. Situation des Archives de l'Etat du Valais (AEV)

Contrairement aux autres services d'archives membres du groupe de travail, les AEV n'ont pas effectué de versement d'archives depuis le système CARI utilisé par le Canton du Valais. Elles disposent toutefois d'une plate-forme de pérennisation d'archivage en fonction depuis 2011 au sein de l'Administration cantonale valaisanne et qui gère aujourd'hui environ 25TB de données.

Les AEV ont également développé, avec l'entreprise Docuteam, le Matterhorn METS Profile¹, utilisé pour la création des paquets d'information (SIP) et l'archivage des données. Ce format est utilisé de manière standard pour tous les versements réalisés sur la plate-forme de pérennisation et d'archivage.

Les résultats de ces différents projets sont mis à profit dans ce document, afin de soutenir les propositions réalisés. Dans le dernier point l'expérience des AEV avec le Matterhorn METS Profile permet d'illustrer certaines possibilités qui pourraient être appliquées pour l'archivage des données issues de ViaCar ou CARI.

¹ <http://www.loc.gov/standards/mets/news112912.html>

2. Formats d'archivage

2.1. Solutions retenues jusqu'ici par les cantons

Pour les projets menés par les cantons membres du groupe de travail, les solutions suivantes ont été appliquées :

Staatsarchiv Zürich (StAZH)

Les données ont été exportées depuis ViaCar sous forme de tables en format TXT. Elles ont été converties au format CSV au moment de l'archivage.

Staatsarchiv Aargau (StAAG)

Les données primaires ont été exportées depuis ViaCar en format CSV, format dans lequel elles sont archivées.

Staatsarchiv Graubünden (StAGR)

Les données primaires ont été exportées depuis CARI en format CSV, format dans lequel elles sont archivées.

Archives de l'Etat du Valais (AEV)

Les AEV ont défini une liste de formats standards qui sont utilisés par la plate-forme de pérennisation et d'archivage, selon les formats d'origine des données ou documents :

<i>Domaine d'utilisation</i>	<i>Formats pérennes</i>
Documents bureautiques standards	PDF/A-1a
Tableaux	CSV
Bases de données relationnelles	SIARD RDB DATA
Images	TIFF 6
Audio	WAVE
Vidéo	DV

2.2. Options pour le format d'archivage

Le choix du format d'archivage dépendra du modèle de données définitivement retenu au moment de l'archivage. Pour mémoire, deux options sont envisagées par la groupe de travail :

1. **Modèle relationnel** : ce modèle implique que chaque table de la base de données soit conservée séparément. Seule la liaison des différents tables permet de rendre compréhensible le lien qu'il existe entre les données.
2. **Modèle dé-normalisé** : les données concernant une entité spécifique (par exemple le permis de conduire) sont conservées ensemble sous forme de tableau. Toutes les données liées à cette entité sont conservées dans ce même tableau qui est compréhensible par lui-même.

Modèle relationnel

Pour ce cas, le format SIARD (eCH-0165)² est la référence au niveau suisse et européen pour l'archivage de bases de données relationnelles. SIARD implique de considérer chaque base de données comme un seul document à archiver. Si un archivage périodique de la base de données est retenu, cela impliquera que chaque versement représentera un document.

Modèle dé-normalisé

Pour ce cas, la conservation des données sous forme de tableaux peut se faire dans le format CSV (*Comma Separated Value*), solution qui a été retenue pour les différents projets (StAZH, StAAG, StAGR).

² <http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165&documentVersion=1.0>

3. Structure et profil du SIP

3.1. Solutions retenues jusqu'ici par les cantons

Pour les projets menés par les cantons membre du groupe de travail, les solutions suivantes ont été appliquées :

Staatsarchiv Zürich (StAZH)

Les StAZH ont utilisé le format eCH-0160³. Le SIP a été généré à l'aide de l'outil Package Handler développé par les Archives fédérales (AFS). Les métadonnées définies par eCH-0160 ont été complétées au moment du versement. Une documentation complémentaire sous format PDF a été préparée pour décrire textuellement le versement et son contenu. Cette documentation est conservée hors du SIP, dans le dossier du versement tenu par les StAZH.

Staatsarchiv Aargau (StAAG)

Les StAAG ont utilisé le format eCH-0160. Le SIP a été généré à l'aide de l'outil Package Handler. Les métadonnées descriptives et techniques ont été incluses directement dans le SIP.

Staatsarchiv Graubünden (StAGR)

Les StAGR n'utilisent pour l'instant pas le profil eCH-0160. Les versements ont été gérés avec l'outil Docupack de Docuteam qui permettrait de prendre en charge le profil eCH-0160, mais les StAGR n'ont pas d'application permettant de générer des versements sous ce format. Les métadonnées complétées concernent essentiellement le versement. Une documentation en format PDF servant à décrire textuellement le versement et son contenu a été intégrée au SIP.

Archives de l'Etat du Valais (AEV)

Les AEV utilisent le Matterhorn METS Profile pour l'archivage électronique. Il sert à générer tous les SIP des versements réalisés sur la plate-forme de pérennisation et d'archivage. Les versements sont générés avec les outils de Docuteam (Docupack et Documill). Les données techniques des versements sont gérées grâce à la récolte des métadonnées issues des systèmes d'origine et gérées dans le Matterhorn METS Profile. Il n'existe pas de documentation annexe aux versements, mais un protocole de versement d'archives électroniques est signé au début de chaque projet afin de définir les conditions générales entre le service producteur et les AEV. Il s'agit d'un document administratif et non technique, qui n'est pas intégré aux versements.

3.2. Options pour les formats de SIP

Il existe aujourd'hui différents profils de SIP, dont plusieurs ont été utilisés par les différents services d'archives. Les deux principalement utilisés en suisse, et parmi les membres du groupe de travail sont :

- eCH-0160, qui est le standard national utilisé par les AFS et plusieurs cantons ;
- le Matterhorn METS Profile, qui est employé par les AEV et différentes institutions en Suisse.

eCH-0160

Formats de fichier

eCH-0160 n'impose aucun format de fichier (voir exigence P_3.2-1)⁴. Ce point doit donc être défini par chaque service d'archives. Il est donc possible de créer des versements de données CSV ou SIARD compatibles avec eCH-0160, comme l'ont fait les StAZH ou les StAAG.

³ <http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0160&documentVersion=1.1>

⁴ les renvois sont fait vers les numéros des exigences concernés dans eCH-0160

Information de pérennisation

Les éléments techniques liés au versement, au système d'origine (ici ViaCar ou CARI) ou aux formats de fichier employés doivent être rassemblés dans un fichier annexe. Ce fichier doit être intégré dans le SIP en créant une structure de classement simple particulière (voir exigence M_4.8-4).

Information de description

Le modèle de données défini par eCH-0160 contient les entités suivantes obligatoires, que doivent renseigner les métadonnées (voir exigence M_4.4-1) :

- provenance
- paquet
- versement
- système de classement
- position du système de classement
- table des matières
- dossier
- fichier
- répertoire

Certaines, telles que « système de classement » ou « position du système de classement » semblent moins adaptées à un versement issu d'une base de données, mais plutôt destinés aux documents issus d'un système GEVER.

Taille maximale des versements

La taille maximale théorique conseillée d'un versement est de 8GB, mais les AFS ont déjà réalisé des SIP dont la taille dépasse les 90 GB.

Matterhorn METS Profile

Formats de fichier

Le Matterhorn METS Profile a été paramétré selon les formats standards présentés au point 2.1. Il intègre donc les formats SIARD et CSV.

Métadonnées techniques et de pérennisation

Les données techniques liées au versement sont récoltées directement au moment de la création du SIP et empaquetées selon le Matterhorn METS Profile (elles sont structurées selon le catalogue PREMIS). Il est également possible d'y enregistrer ensuite toutes les modifications ou mises à jour qui ont lieu lorsque les données sont stockées sur la plateforme d'archivage (migration de format, intervention des archivistes, etc.).

Il est possible d'imaginer récolter et décrire ici les métadonnées liées à la gestion de la base de données d'origine (acteurs, droits d'accès, audits, etc.). Cela permettrait une plus grande richesse dans la description des données.

Métadonnées descriptives

Les métadonnées descriptives sont organisées selon les standards de description du Conseil international des archives (ICA) :

- Documents et données : ISAD(G), encodé en EAD ;
- Producteur des documents et données : ISAAR(CPF), encodé en EAC ;
- Fonctions exercées pour la production des documents et données : ISDF, encodé en proto-EAF (en cours de développement par l'ICA).

Il est donc possible de récolter les informations concernant non seulement les documents et données, mais également leur producteur et les fonctions exercées pour la production des données (pour autant qu'elles existent).

Taille maximale des versements

La taille maximale paramétrée est de 8 TB. Le versement le plus volumineux réalisé jusqu'ici est de 500 GB.

Ces deux options restent ouvertes et sont à discuter, selon les pratiques des différents services d'archives.

4. Organisation des versements aux Archives

4.1. Responsabilités en matière d'archivage

L'organisation des versements aux Archives dépendra beaucoup de la variante retenue en matière d'archivage et de gestion des données. Trois options restent ouvertes :

- Un archivage selon le modèle de données (*Datenmodell*) défini par le groupe de travail, réalisé par les cantons (situation existante jusqu'ici) ;
- Un archivage selon le modèle de données (*Datenmodell*) défini par le groupe de travail réalisé par les AFS depuis la future base de données fédérales IVZ/SIAC ;
- Un archivage centralisé par les AFS pour les données gérées dans IVZ/SIAC et un archivage complémentaire par les cantons pour les données qui sont uniquement gérées au niveau cantonal.

La situation n'est pas encore très claire sur ces points et il faut tenir compte du fait que le projet IVZ/SIAC est actuellement bloqué. Sa mise en œuvre est encore incertaine.

4.2. Organisation technique des versements

Les aspects techniques des versements dépendront bien sûr de l'organisation choisie, mais deux variantes peuvent être envisagées :

- Une extraction « manuelle » des données par les services de circulation routière selon la périodicité définie, puis la préparation du SIP et le versement aux Archives. C'est ce qui a été appliqué dans les projets menés par les cantons membres du groupe de travail.
- Une extraction automatisée directement depuis les systèmes ViaCar ou CARI.

Les AEV testent actuellement cette seconde option pour l'archivage des données depuis le système ECM (GEVER) de l'Administration cantonale valaisanne. L'avantage d'une telle option est de permettre un archivage en grande partie automatisé, mais elle nécessite :

1. De définir un « élément déclencheur » dans le système d'origine (clôture d'un dossier, durée de conservation, fin d'une année, etc.).
2. De définir l'interface technique pour l'extraction et le transfert des données (voir le Work Package « Schnittstelle »).

Florian Vionnet
Archiviste-records manager