

# Les métadonnées dans les paquets d'information SIP et AIP. De leur création à leur gestion<sup>1</sup>

Brigitte Kalbermatten

## Introduction

Avec l'avènement de la société numérique, une nouvelle forme d'information doit être pérennisée par les archivistes, bibliothécaires et autres professionnels de la conservation du patrimoine culturel. Il s'agit des données électroniques (des documents Word ou Excel, des sites internet, des films, des bases de données, etc.) qui se caractérisent par leur fragilité et leur durée de vue éphémère.<sup>2</sup> L'archivage numérique se veut un moyen de conservation et de préservation à long terme de ces données dématérialisées. Il visera donc à garantir leur longévité, leur intégrité, leur authenticité et leur accessibilité. Néanmoins, ce secteur d'activité reste jeune, avec relativement peu de retour d'expériences.

Depuis presque une décennie, l'archivage numérique se base sur le modèle de référence pour un Système ouvert d'archivage d'information (ci-après OAIS),<sup>3</sup> qui a été publié en 2002 par le Comité Consultatif pour les Systèmes de Données Spatiales. Cette norme internationale propose une approche globale et conceptuelle de la problématique posée par l'archivage numérique. Parmi les concepts qu'elle définit, le paquet d'information nous intéresse tout particulièrement. Il s'agit de regrouper les données à pérenniser avec les métadonnées les documentant. C'est ce second élément qui va nous intéresser dans cette contribution.

Au fait, à quoi faisons-nous référence quand nous parlons de métadonnées? OAIS la définit comme étant une « data about other data ».<sup>4</sup> Pour une meilleure compréhension, il faut la compléter avec l'explication donnée par la *National Science*

---

1 Cet article se base sur le travail de diplôme pour le Master of Advanced Studies in Archival, Library and Information Science: Kalbermatten, Brigitte: Les métadonnées dans les paquets d'information SIP et AIP. De leur création à leur gestion. Berne 2012.

2 Comme risques majeurs, il faut citer la fragilité du support, dans le sens qu'un seul bit endommagé empêche la lecture d'un fichier, et l'obsolescence du matériel ainsi que des logiciels nécessaires à sa création et à sa lecture.

3 L'abréviation OAIS vient de l'anglais *Open Archival Information System*. La dernière version du modèle a été publiée en juin 2012. Elle est disponible à l'adresse internet suivante: <http://public.ccsds.org/publications/archive/650x0m2.pdf> (consultée le 9 juin 2013).

4 Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System, 2012, 1–13.

*Foundation*: «Metadata summarize data content, context, structure, interrelationships, and provenance (information on history and origins). They add relevance and purpose to data, and enable the identification of similar data in different data collections.»<sup>5</sup>

Problématique et méthode

Toute institution d'archives qui projette un système d'archivage numérique s'interroge sur la question des métadonnées. Avec le modèle de référence OAIS, elle dispose d'informations générales et conceptuelles sur le sujet, mais pas de solutions toutes faites. Dès lors, une part de leur travail de réflexion consiste à déterminer ce qu'elle souhaite avoir comme métadonnées alors que l'autre part se focalise sur l'étude de profils de métadonnées<sup>6</sup> déjà existants.

En Suisse, nous pouvons dénombrer actuellement deux profils de métadonnées qui ont fait, par ailleurs, l'objet d'une standardisation : l'un est proposé par les Archives fédérales suisses et le Centre de coordination pour l'archivage à long terme de documents électroniques (CECO), et l'autre par les Archives de l'Etat du Valais et *Docuteam GmbH*.

Cette contribution se propose de présenter et comparer ces deux profils pour déterminer leur interprétation du modèle de référence OAIS ainsi que leur principal avantage et leurs désavantages. Pour le Matterhorn METS Profile, ces désavantages tendent plutôt à être de nouveaux défis ou activités que les archivistes responsables de l'archivage numérique se doivent de prendre en compte et gérer sur le long terme pour remplir leur mission de pérennisation des données électroniques. Autrement dit, ils influent directement sur le métier de l'archiviste et sur la conception du cycle de vie des documents. Le champ d'étude couvre les métadonnées contenues dans les paquets d'informations à verser (abrégé en SIP) et ceux à archiver (désormais AIP),<sup>7</sup> qui sont quasiment identiques dans les deux cas.

## Les métadonnées selon OAIS

Etant un modèle conceptuel, OAIS identifie un certain nombre de catégories de métadonnées, mais sans indiquer leur contenu exact, qui aurait pu être sous la forme d'un dictionnaire de données.

5 National Science Foundation : Cyberinfrastructure Vision for 21st Century Discovery, Arlington 2007, 22.

6 Un profil de métadonnées indique les différentes métadonnées employées et regroupées en catégories ainsi que leurs relations.

7 Ces deux catégories proviennent d'OAIS qui définit trois types de paquets d'information : *Submission Information Package (SIP)* pour les données à verser, l'*Archival Information Package (AIP)* pour les conserver et *Dissemination Information Package (DIP)* pour les communiquer.

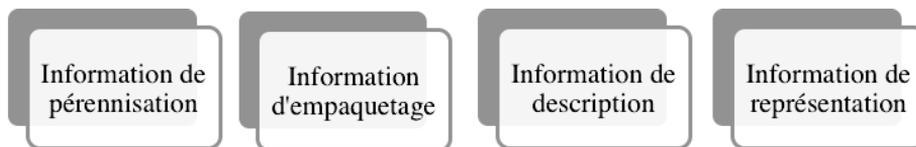


Figure 1 : Catégories de métadonnées d'après OAIS

L'information de pérennisation (*Preservation Description Information*) constitue l'élément principal qui permet à une institution d'archives d'assurer sa mission de conservation à long terme des données électroniques. Elle se compose de cinq sous-catégories : l'information de contexte qui permet à un document numérique d'être rattaché à un dossier,<sup>8</sup> l'information de provenance qui retrace les événements l'affectant, l'information d'identification qui permet de l'identifier de manière unique et pérenne, l'information d'intégrité qui fournit des outils de validation ou de vérification afin de lui éviter toute altération accidentelle, et l'information sur les droits d'accès qui peuvent limiter sa consultation pour des raisons légales.

L'information d'empaquetage (*Packaging Information*) met en relation les composants du paquet, autant les données électroniques entre elles que les métadonnées associées à chacune.

L'information de description (*Descriptive Information*) est constituée des catégories précédentes. Elle permet de rechercher, de trouver et d'analyser les données électroniques pérennisées.

L'information de représentation (*Representation Information*) doit permettre de transformer la succession de bits (0 et 1), qui forme tout document numérique, en un contenu intelligible pour l'homme. Elle se compose de l'information de structure (description du format) et de celle sémantique (par exemple, l'indication de la langue d'un texte). Sa conservation est rendue obligatoire par le fait que sa durée de vie est relativement courte.<sup>9</sup>

### Standard eCH-0160 : Interface de versements d'archives (SIP)<sup>10</sup>

A l'origine de ce standard se trouve le profil de métadonnées des Archives fédérales suisses qui a été retravaillé en collaboration avec le Centre de coordination pour l'archivage à long terme de documents électroniques dans le but de permettre son

8 Autrement dit, elle documente la création du document (auteur, date, etc.).

9 Prenons comme exemple la suite bureautique de Microsoft (Office) dont une nouvelle version sort en moyenne tous les trois ans.

10 Sa documentation est disponible sur la page internet suivante : [www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0160](http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0160) (consultée le 9 juin 2013).

utilisation hors de l'administration fédérale. L'association eCH l'a accepté comme standard national en novembre 2012. Cet organisme de standardisation adopte des normes pour la cyberadministration en Suisse.

Ce profil a été conçu pour des versements de documents issus de logiciels dédiés à la gestion électronique des documents (paquet d'informations de type GEVER<sup>11</sup>) ou de systèmes de fichiers et de bases de données relationnelles (type FILES).

#### Description du profil de métadonnées

La documentation se compose d'une spécification (qui définit comment un versement numérique peut et doit se faire), d'un dictionnaire de données et d'un schéma XSD.<sup>12</sup> Comme la lecture de fichiers encodés en XML n'est pas aisée, nous ne rentrons pas dans les détails de ce dernier document qui constitue, en fait, une vue réelle de la structure d'un paquet d'informations.

La spécification propose un modèle conceptuel de données qui se compose de treize classes UML<sup>13</sup> non détaillées et de leurs relations. Dix de ces treize classes constituent des entités explicitement présentes dans le SIP et décrites dans le dictionnaire de données. D'après OAIS, toutes peuvent être considérées comme des informations de description.

La classe intitulée «Paquet» contient des informations d'empaquetage qui permettent de comprendre la structure du paquet, de l'identifier et de documenter très sommairement les événements l'ayant affecté.<sup>14</sup>

L'entité «Versement» informe sur le contexte, en particulier sur la provenance, et sur les droits d'accès défini d'un commun accord entre le service versant et les archives pour le SIP entier.

Les informations contenues dans «Provenance» documentent le contexte et la provenance en proposant des champs de métadonnées sur le producteur des documents (nom, période d'existence et histoire).

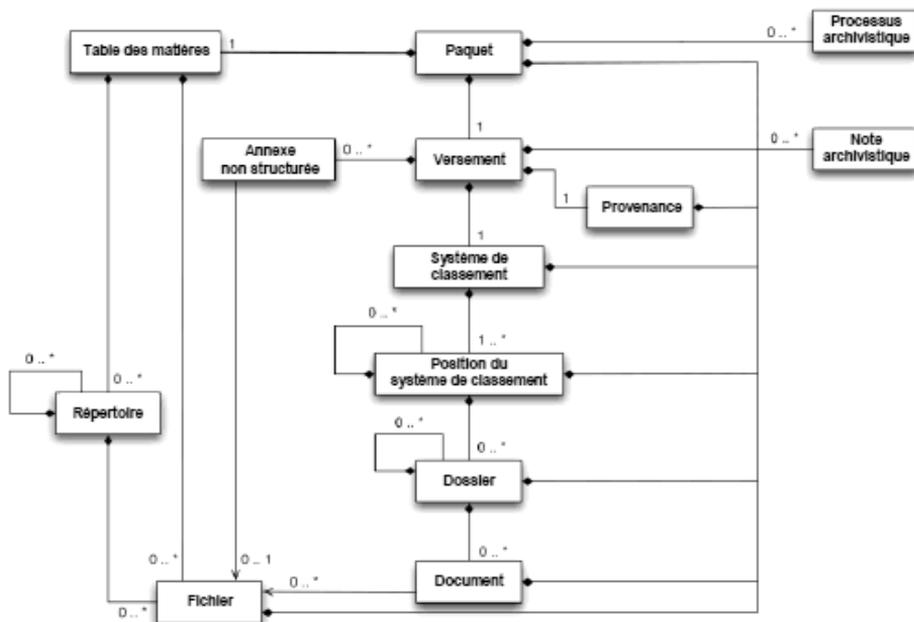
Les informations de contexte sont encore complétées par l'entité «Système de classement» qui se concentre sur le logiciel de gestion des documents (si existant), une base de données ou toute autre source d'informations à pérenniser.

11 GEVER est l'acronyme pour la gestion électronique des documents au sein de l'administration fédérale.

12 Les métadonnées sont généralement encodées en XML (*Extensible Markup Language*). Leur choix et leur structure sont définis dans un schéma nommé XSD, soit XML Schema Definition, qui peut être sous la forme d'un ou plusieurs documents XML.

13 Unified Modeling Language (UML) : il s'agit d'un langage de modélisation graphique au moyen de pictogrammes qui est très utilisé dans les projets logiciels.

14 Un champ de métadonnées («Version») a été prévu pour indiquer la version du paquet, autrement dit s'il a déjà été migré ou non. Voir Groupe spécialisé Archivage numérique : Interface de versements d'archives (SIP) – Data Dictionary, Zurich 2012, 1.



**Figure 2 : Modèle conceptuel des données**  
(Groupe spécialisé Archivage numérique : Interface  
de versements d'archives [SIP], Zurich 2012, 18)

Les trois classes suivantes, « Position du système de classement »,<sup>15</sup> « Dossier » et « Document », forment la structure du plan de classement du fonds d'archives, ou tout du moins la partie concernée par le versement. Elles renseignent sur l'identification, le contexte et les droits d'accès, mais à leur niveau respectif. Aucun renseignement relatif aux caractéristiques techniques des fichiers n'est indiqué ici. Ce genre d'informations se trouve dans l'entité « Fichier ».

En listant tous les fichiers et les dossiers constituant le SIP ou l'AIP, la « Table des matières » appartient à la catégorie définie par OAIIS pour les informations d'emballage.

L'entité « Répertoire », qui ne doit pas être confondue avec celle du « Dossier », concerne essentiellement l'identification du contenant physique du fichier versé.

Quant à la classe « Fichier », elle reprend les mêmes informations que la précédente, mais en les complétant avec des métadonnées relatives aux sommes de contrôle. Ces dernières représentent un mécanisme de vérification de l'intégrité des données conservées.

15 Cette notion vient de l'allemand *Ordnungssystemposition* et du monde de la gestion des documents plutôt que de l'archivistique. En français, elle équivaut à la série.

«Processus archivistique» documente les événements qui touchent principalement la vie du paquet, de sa réception aux archives (contrôles de qualité du versement : formats, présence ou non de virus, etc.) aux mesures de préservation (par exemple, la migration).

La classe «Annexe non structurée» décrit des fichiers ajoutés au paquet d'informations en tant que renseignements supplémentaires. Selon la spécification, elle ne devrait plus être employée et sera peut-être même supprimée dans une prochaine version du standard.<sup>16</sup>

Enfin, la «Note archivistique» permet aux archivistes d'annoter un ou plusieurs objets lors des contrôles manuels du versement et du processus d'archivage dans le magasin numérique.

En observant le dictionnaire de données, nous constatons que les métadonnées du SIP sont quasiment identiques à celles de l'AIP. En effet, le document contient déjà des champs de métadonnées relatifs à la gestion des AIPs, à savoir leurs différentes identifications dans l'entité «Paquet» et la consignation des événements dans «Processus archivistique».

#### Adéquation avec OAIS

Le modèle de référence OAIS définit des catégories d'information qui concernent directement ou indirectement un ou plusieurs objets d'information à pérenniser et le paquet d'informations lui-même. Le niveau de détail est laissé au choix des institutions responsables de l'archivage numérique, car les concepteurs d'OAIS le considèrent déjà probablement comme une étape de la mise en application d'un tel système.

Pour le standard eCH-0160, la granularité reste à un niveau élevé, en particulier pour les informations d'intégrité et de droits d'accès. Certes, les sommes de contrôle sont calculées et vérifiables pour les fichiers, mais la traçabilité des événements et les délais de protection sont indiquées respectivement au niveau du paquet et dans les entités «Position du système de classement», «Dossier» et «Document».

En outre, aucune information de représentation n'est disponible dans ce profil de métadonnées. Aucun champ de métadonnées n'a été prévu pour indiquer le format. Un moyen d'introduire ce renseignement serait d'utiliser le champ de métadonnées «Propriétés» en définissant la paire d'attributs comme suit : propriété = format et valeur = nom du format.

A long terme, il pourrait être pertinent de définir un champ expressément dédié aux informations de représentation. Dans l'idéal, le nom du format et un lien (non dynamique, sous la forme d'un identifiant) vers un référentiel regroupant les

<sup>16</sup> Groupe spécialisé Archivage numérique : Interface de versements d'archives (SIP), Zurich 2012, 16.

informations techniques sur les formats adaptés à l'archivage numérique devraient être introduits dans ce profil de métadonnées.

#### Avantage et désavantages

Comme le profil de métadonnées se décline en deux types (GEVER ou FILES), l'évaluation des métadonnées n'est pas une tâche requise. Ainsi, elle n'implique pas l'engagement de ressources humaines et financières qui peuvent donc être utilisées pour d'autres activités et projets du centre d'archives. Découlant de cette constatation, ce profil de métadonnées tend à permettre une plus grande automatisation du processus de versement des documents, sans devoir systématiquement analyser et évaluer les métadonnées générées par les divers logiciels à l'origine des données électroniques à pérenniser. Cet avantage s'adapte bien à l'administration fédérale, dont le nombre de logiciels est relativement limité. Par exemple, ce sont Fabasoft et GEVER Office qui sont principalement utilisés à l'heure actuelle pour la gestion des documents.

L'inconvénient majeur de ce standard se rapporte à son orientation qui est très, voire trop, tournée sur les besoins de l'administration fédérale. Il peut représenter un frein certain à son utilisation dans d'autres organismes publics (cantons ou communes) et privés (entreprises). A titre d'exemple, mentionnons la présence de notions spécifiques, comme GEVER, ou de contraintes issues de l'environnement propre aux Archives fédérales suisses, telle l'absence d'un champ de métadonnée pour la cote. Le premier exemple se réfère au projet de mise en œuvre du Records Management au sein de l'administration fédérale. Quant au second, il tend à montrer que le profil de métadonnées n'est pas suffisamment neutre et générique pour d'autres institutions : le champ cote peut être très utile pour des institutions faisant *de la numérisation* de documents qu'elles conservent et qui portent déjà une cote. Certes, il est possible d'utiliser le champ de métadonnée « Données supplémentaires », mais il s'agit d'une solution « fourre-tout ». Par analogie à l'évaluation des plans de classement, les positions « Divers » contiennent tout ce qui ne peut pas être classé ailleurs et ne sont pas conservées sur le long terme. Les bonnes pratiques tendent à éviter le plus possible leur utilisation, ce qui devrait aussi se faire pour les métadonnées et le champ « Données supplémentaires ».

L'absence d'information de représentation peut poser un problème pour la conservation des données électroniques et pour les mesures de préservation. Le paquet d'informations n'est dès lors plus indépendant de l'environnement logiciel. En effet, cette information peut exister dans le logiciel de gestion de la plateforme d'archivage, mais elle n'est, dans ce cas, pas directement rattachée au fichier concerné.

Ne s'appuyant pas sur des standards de métadonnées connus internationalement, ce profil de métadonnées peut engendrer des coûts supplémentaires lors de

l'acquisition du logiciel de gestion de la plateforme d'archivage numérique, dont les fournisseurs sont plutôt étrangers à la Suisse. L'un des tests proposé par ce genre d'application consiste à garantir la conformité du fichier contenant les métadonnées. Il faut dès lors le personnaliser entièrement selon ce profil de métadonnées.

### **Matterhorn METS<sup>17</sup> Profile**

Ce profil de métadonnées a été élaboré en collaboration par les Archives de l'Etat du Valais et par l'entreprise Docuteam GmbH. Il s'appuie sur les deux standards internationaux et principaux de l'archivage numérique : METS et PREMIS,<sup>18</sup> qui sont créés et maintenus par la Bibliothèque du Congrès (Etats-Unis). De plus, il combine les trois normes du Conseil international des archives : ISAD(G)<sup>19</sup> pour la description des documents d'archives, ISAAR(CPF)<sup>20</sup> pour les producteurs et ISDF<sup>21</sup> pour les fonctions. En novembre 2012, le profil de métadonnées a été publié par la Bibliothèque du Congrès sur son site internet. Il est devenu «the first profile that describes the use of EAD within METS in any detail»,<sup>22</sup> c'est-à-dire qu'il est le premier profil à contenir la description des documents d'archives sous la forme de la description archivistique encodée (*Encoded Archival Description*, EAD).

Le Matterhorn METS Profile est un profil générique qui s'adapte à différentes utilisations. Il répond à une exigence émise lors de sa création, puisque la plateforme d'archivage des Archives de l'Etat du Valais sert également à la Médiathèque Valais et aux Musées cantonaux.<sup>23</sup> Ce profil représente donc le premier niveau d'une hiérarchie en comportant quatre au total.

D'abord, le profil générique contient les métadonnées communes à chaque paquet d'informations.

Vient ensuite le profil utilisé par chaque institution. La différence majeure à ce niveau correspond au standard utilisé pour la description : il s'agit de l'EAD pour les Archives de l'Etat du Valais et du Dublin Core pour la Médiathèque Valais et les Musées cantonaux.

17 METS est l'abréviation de *Metadata Encoding and Transmission Standard*. Voir le site officiel disponible à l'adresse suivante : [www.loc.gov/standards/mets/](http://www.loc.gov/standards/mets/) (consultée le 16 juin 2013).

18 PREMIS signifie *Preservation Metadata: Implementation Strategies*. Pour plus d'informations, voir [www.loc.gov/standards/premis/](http://www.loc.gov/standards/premis/) (consulté le 16 juin 2013).

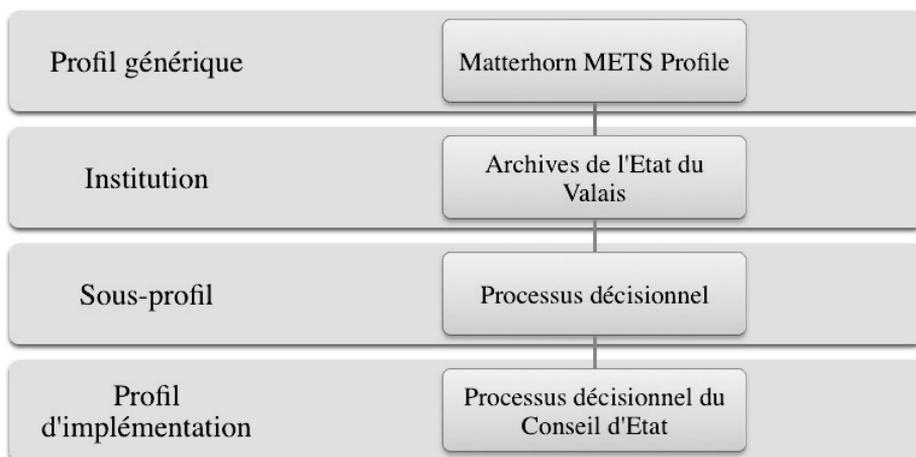
19 Norme générale et internationale de description archivistique ISAD(G), publiée en 2000.

20 Norme internationale sur les notices d'autorité archivistiques relatives aux collectivités, aux personnes et aux familles ISAAR(CPF), publiée en 1996.

21 Norme internationale pour la description des fonctions ISDF, publiée en 2007.

22 [www.loc.gov/standards/mets/news112912.html](http://www.loc.gov/standards/mets/news112912.html) (consulté le 16 juin 2013).

23 Autrement dit, la plateforme d'archivage appartient au Service de la culture, mais est administrée par les Archives de l'Etat du Valais.



**Figure 3 : Structure en quatre niveaux d'un SIP contenant des documents créés lors du processus de décision du Conseil d'Etat**

Le niveau suivant se détermine par le processus à l'origine des documents (exemples : processus décisionnel, de numérisation, etc.).

Enfin, le profil d'implémentation est constitué des métadonnées définitives du paquet d'informations versé et archivé. Il est adapté après une évaluation des métadonnées générées par le logiciel à l'origine des documents.

La figure 3 montre un exemple concret d'un paquet numérique contenant des documents créés lors du processus décisionnel du Conseil d'Etat.

#### Description du profil de métadonnées

La documentation du Matterhorn METS Profile se résume à l'heure actuelle au schéma XML disponible sur le site internet de la Bibliothèque du Congrès. Sur demande aux Archives de l'Etat du Valais ou à *Docuteam GmbH*, une carte de métadonnées est disponible.

Comme auparavant, le schéma XML est laissé de côté afin de pouvoir se concentrer sur sa représentation graphique qu'est la carte de métadonnées. Ce document est construit à l'aide de classes UML détaillées : les attributs, sous la forme de nom des métadonnées, sont mentionnés. Bien évidemment, les relations entre les différentes classes sont également indiquées, mais ne seront pas toutes commentées dans les pages qui suivent.

Le standard METS sert de container aux métadonnées. Sa structure se compose de cinq parties : l'en-tête METS, la section des fichiers et la carte de structure dont le contenu correspond à l'information d'empaquetage d'OAIS, la section des

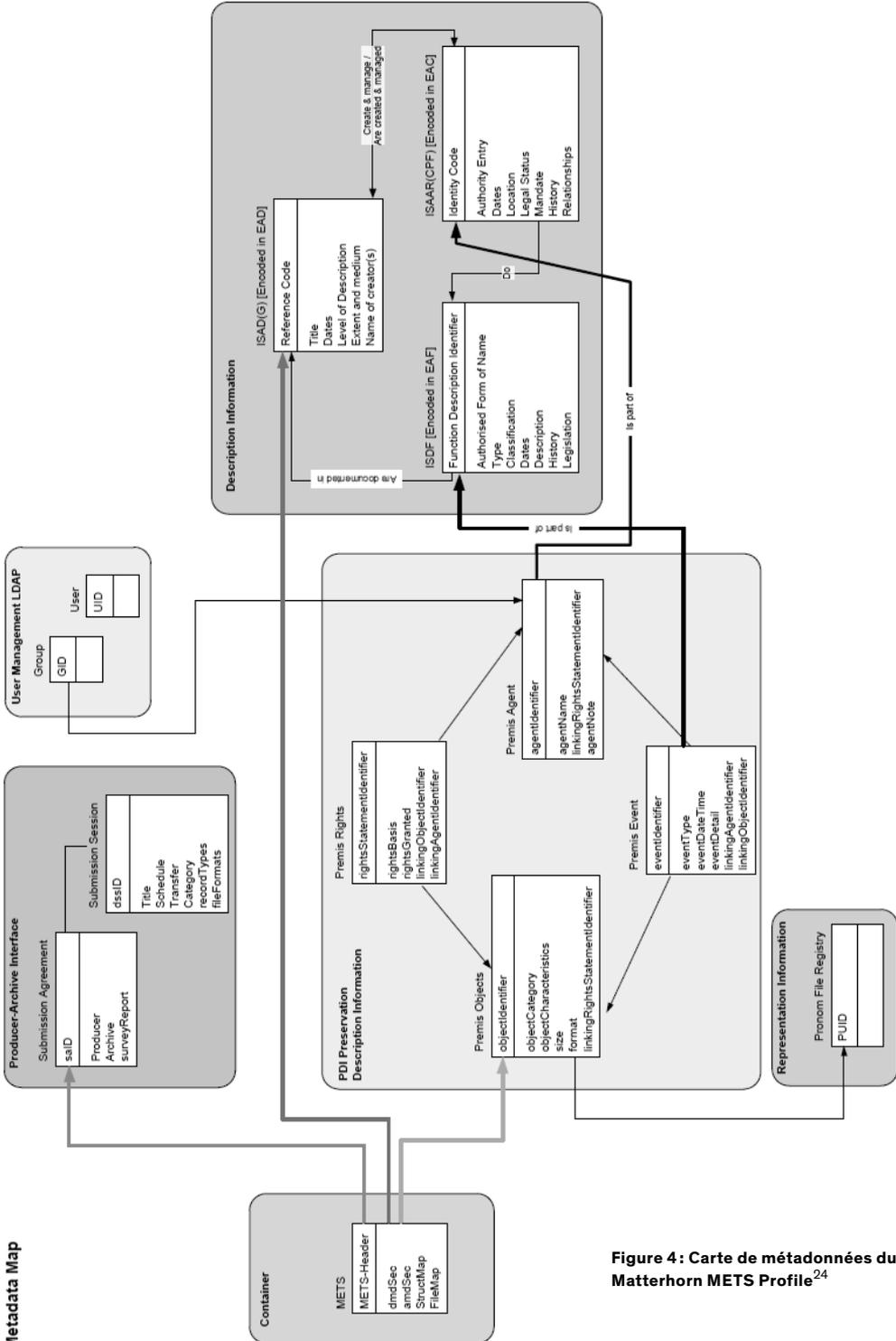


Figure 4 : Carte de métadonnées du Matterhorn METS Profile<sup>24</sup>

métadonnées descriptives qui représente l'information de description et la section des métadonnées administratives où se trouve l'information de pérennisation.

Graphiquement parlant, ce container METS est lié à l'interface entre le producteur et les archives. Cette information n'est pas une catégorie d'informations en tant que telle, mais un outil nécessaire pour régler les versements de documents. Il s'agit du protocole de versement (*submission agreement*) qui documente la provenance des données électroniques et les modalités du versement. Dans ce document est, entre autres, défini le profil de métadonnées à employer.

L'information de pérennisation (le rectangle au centre de la figure 4) est encodée en PREMIS. Ce standard international propose un dictionnaire de données nécessaires au bon fonctionnement d'une infrastructure pour l'archivage numérique. Sa conception envisage que les métadonnées soient complétées automatiquement par les logiciels liés à l'archivage numérique. PREMIS se compose de cinq entités, dont quatre sont ici présentes. L'entité «Objet» permet de documenter techniquement le (ou les) fichier(s) pérennisé(s) avec, par exemple, le format, la taille ou un lien vers les droits d'accès correspondants. Ces derniers se trouvent dans l'entité «Droits» qui contient, sous la forme de métadonnées ou de liens, toutes les informations relatives aux droits d'accès, soit quels droits sur quels objets possède tel groupe d'utilisateurs ou agents. Cette notion est justement le titre de la troisième entité de PREMIS. Elle donne le nom des agents et des liens, dont un qui pointe vers les événements que l'agent peut effectuer. Les renseignements sur les événements forment la dernière entité, intitulée «Événement». Il s'agit essentiellement du type et de la date de l'événement, avec quelques liens supplémentaires.

L'entité «Agent» de PREMIS est liée au «User Management LDAP», qui doit gérer à travers le temps les groupes d'utilisateurs au sein de l'administration cantonale. Autrement dit, «tant qu'un objet est soumis à un délai de protection, tout changement dans la composition des agents dans le système de gestion des documents se répercute automatiquement sur la plateforme d'archivage».<sup>25</sup>

L'information de représentation est constituée de la mention du registre des formats PRONOM des Archives nationales d'Angleterre<sup>26</sup> et de l'identifiant du format mentionné dans l'information de pérennisation.

Comme mentionné plus haut, l'information de description combine plusieurs normes de description, chacune encodée dans son format spécifique. Les documents sont décrits avec ISAD(G). L'une des métadonnées, le nom du (ou des) créateur(s),

24 Publiée lors du forum du 17 juin 2013 des archivistes genevois : [www.forumdesarchivistes.ch/2013/06/24/les-normes-de-description-archivistique-un-potentiel-inconnu/](http://www.forumdesarchivistes.ch/2013/06/24/les-normes-de-description-archivistique-un-potentiel-inconnu/) (consultée le 30 juin 2013).

25 Kalbermatten, Brigitte : Les métadonnées dans les paquets d'information SIP et AIP. De leur création à leur gestion. Berne 2012, 54.

26 Registre disponible à l'adresse suivante : [www.nationalarchives.gov.uk/PRONOM/Default.aspx](http://www.nationalarchives.gov.uk/PRONOM/Default.aspx) (consultée le 16 juin 2013).

permet de faire le lien avec ISAAR(CPF) qui sert à décrire leur producteur. Les fonctions de ce dernier sont explicitées à l'aide d'ISDF et renvoient elles-mêmes aux documents. Cette relation entre ces trois standards pourrait être décrite autrement : le document est créé ou reçu par le producteur dans l'exercice de ses fonctions.

#### Adéquation avec OAIS

Les quatre catégories d'information énoncées par OAIS – information de pérennisation, d'empaquetage, de description et de représentation – sont présentes dans le Matterhorn METS Profile. Elles sont clairement délimitées et visibles, car la terminologie est justement issue de ce modèle de référence, sauf pour l'information d'empaquetage où les concepteurs préfèrent la notion de container.

#### Avantage et désavantages

L'utilisation du standard METS apporte un avantage indéniable au Matterhorn METS Profile, puisqu'il est utilisable dans un grand nombre de logiciels de gestion de plateforme d'archivage numérique. De par cette présence, il participe à la garantie de l'interopérabilité entre divers systèmes d'archivage numérique. En outre, l'utilisation de METS n'amène pas de contraintes qui empêcheraient ou gêneraient l'utilisation de ce profil par une bibliothèque. En effet, Dublin Core peut être utilisé avec METS.

L'évaluation des documents papier implique des ressources en temps, financières et personnelles que toute institution conservant des archives investit. Seulement, elle ne concerne pas les métadonnées, car ces dernières existent en nombre restreint. Au contraire, les possibilités de documenter tout le cycle de vie des données ont fortement augmenté avec l'avènement de l'informatique dans notre quotidien. Il est dès lors devenu nécessaire d'effectuer également une évaluation des métadonnées qui sous-entend une analyse du processus dématérialisé à l'origine des documents (lorsqu'il existe). Une nouvelle tâche est ainsi apparue dans le cahier des charges des archivistes. A terme, elle conduira à une plus grande richesse et qualité des métadonnées qui sera à la disposition de la communauté d'utilisateurs des fonds d'archives. Ainsi que nous le voyons, ce premier « désavantage » représente en fait une évolution du métier de l'archiviste.

Comme le Matterhorn METS Profile utilise PREMIS pour l'information de pérennisation, une difficulté peut apparaître quand les données électroniques à conserver ne proviennent pas de logiciels de gestion des documents. Rappelons que ce standard préconise une saisie automatique des métadonnées et aucune intervention directe de l'homme. Néanmoins, la plupart des activités informatiques<sup>27</sup> sont ins-

<sup>27</sup> Par exemple, le lancement et l'activité du système d'exploitation Windows 7 (édition familiale) génère un certain nombre de fichiers log, qui sont consultables par l'observateur d'événements.

crites dans des journaux d'événements, communément appelés fichiers log (à savoir un simple fichier texte). Prenons comme exemple les documents numérisés pour lesquels le processus de numérisation doit être documenté. Une solution possible consiste à programmer un outil qui extrairait les informations requises depuis les journaux d'événements créés lors de la numérisation et les inscrirait dans les métadonnées du paquet numérique correspondant. C'est cette approche que se proposent d'utiliser les Archives de l'Etat du Valais pour ce genre de cas.

Le Matterhorn METS Profile contient des métadonnées sur les droits d'accès relatifs aux documents encore soumis à un délai de protection. Cette affirmation sous-entend que des dossiers clos, mais encore utilisés par le producteur, sont versés aux archives. Autrement dit, les concepteurs du profil préfèrent la conception d'un cycle de vie continu des documents à la théorie des trois âges,<sup>28</sup> puisqu'elle répond mieux aux besoins actuels en matière de droits d'accès. Cette nouvelle approche du cycle de vie implique un lien dynamique avec l'application qui gère le personnel de l'administration. Les Archives de l'Etat du Valais font l'hypothèse que cette base de données est et sera mise à jour régulièrement (dès que des mutations au sein du personnel ont lieu) par le Service des ressources humaines. Cette situation entraîne le constat suivant : le lien entre les deux applications doit être maintenu actif tant que des documents conservés sur la plateforme de pérennisation et d'archivage seront soumis à un délai de protection et utilisés par le service versant. Elle ne concerne pas l'utilisation des archives par des tiers, dont les droits sont gérés par l'application scopeArchiv. Au final, ce « désavantage » peut différer d'une institution à l'autre et, dans certains cas, constituer un changement de paradigme.

## Conclusion

Chacun des profils étudiés propose son interprétation du modèle de référence OAIS. Le standard eCH-0160 se détache d'OAIS et reste à un niveau de granularité assez haut quant à la majorité des informations de pérennisation. Le Matterhorn METS Profile utilise clairement les quatre catégories d'informations et ce à niveau de détails différent selon les besoins. En effet, la description d'un ouvrage, numérisé ou non, reste relativement sommaire alors que pour une décision du Conseil d'Etat, nous avons des informations précises sur le processus de sa création et sur son contenu.

---

28 En anglais, il s'agit du *Records Continuum* qui abolit les frontières entre les dossiers courants, les archives intermédiaires et les archives définitives que Théodore Schellenberg a introduites avec sa théorie des trois âges en 1956. Cette nouvelle théorie correspond mieux aux besoins de la conservation des documents numériques. Voir Atherton, Jay : *From Life Cycle to Continuum : Some Thoughts on the Records Management – Archives Relationship*. In *Archivaria* 21 (1985–1986), 43–51.

La caractéristique la plus intéressante à souligner vient du Matterhorn METS Profile : une structure en plusieurs niveaux. Cet état de fait permet son utilisation en dehors du domaine des archives. N'oublions pas que les archivistes ne sont pas les seuls à être concernés par l'archivage numérique. Les bibliothécaires, les responsables des musées, les documentalistes, etc. sont également impliqués. Leurs besoins diffèrent de ceux de l'archiviste, puisqu'ils ne gèrent que la valeur d'information de leurs documents. Depuis novembre 2012, ils ont à leur disposition un profil de métadonnées dans lequel ils peuvent utiliser leurs propres formats de description des documents.

Comme nous avons pu le voir, les deux profils de métadonnées ont leur point fort et leurs inconvénients ou impacts sur le travail de l'archiviste, mais ils ont un trait commun : leur jeunesse. Des échanges d'expérience et de la veille technologique sont nécessaires pour leur maturité. Ainsi, autant le standard eCH-0160 que le Matterhorn METS Profile évolueront dans le temps afin de répondre aux nouveaux besoins, tant technologiques pour continuer à garantir la pérennisation des données électroniques qu'intellectuels pour répondre aux attentes de la communauté d'utilisateurs.