

## Spécifications cadres pour l'archivage numérique (projet bentō)

### Sommaire

1	Introduction.....	1
2	AIP.....	2
3	SIP.....	3
4	Information de Description .....	6
5	DIP.....	7
6	Annexes.....	9

### 1 Introduction

Le présent document – principal résultat du projet *bentō* – comprend des spécifications et des frameworks pour les quatre interfaces centrales de l'archivage numérique selon la terminologie de l'OAIS<sup>1</sup>: AIP, SIP, DIP et *Information de Description (Descriptive Information)*. Il s'agit d'une concrétisation des *Exigences de base en matière d'archivage électronique* que la commission de surveillance du CECO a approuvées en septembre 2009 pour servir de fil rouge et de feuille de route<sup>2</sup>.

Du point de vue du modèle OAIS, *bentō* comprend les domaines suivants: *Entrées (Ingest)*, *Gestion de données (Data Management)*, *Stockage (Archival Storage)* et *Accès (Access)*. Nous précisons comme suit le contenu de ces domaines dans le contexte de *bentō*:

- Sur l'interface *Entrées* vers le producteur de données, l'on vérifie que le versement de données se fait sous forme d'un SIP défini, enrichi des métadonnées nécessaires pour l'archivage. Le SIP est ici l'interface avec le producteur de données et est le modèle de données pour les données versées.
- A partir du SIP, le processus *Entrées* génère un ou plusieurs AIP pour la mémorisation à long terme. L'AIP est ici le modèle de données pour une unité archivistique regroupée autonome.
- C'est également dans le processus *Entrées* que l'*Information de Description* est générée pour la *Gestion de données*. Un service de *Gestion de données* est déjà disponible dans toutes les Archives, sous la forme d'un système d'information ou d'un instrument électronique de recherche.
- Le *Stockage* accueille les AIP générés et assure la *Bitstream Preservation*. Le *Stockage* est déjà décrit en tant que service dans le projet CECO *arcun*<sup>3</sup>, et

nous ne donnerons pas d'autres détails à ce sujet dans le cadre de *bentō*.

- Le processus *Accès* permet un accès combiné aux AIP et à l'*Information de Description*. Pour l'utilisation, l'AIP est transformé et converti en un DIP, de manière similaire au processus *Entrées*.

Les domaines de fonctions OAIS *Planification de la pérennisation (Preservation Planning)* et *Administration* restent en dehors du cadre de *bentō*. Dans le modèle OAIS, ils ne sont qu'indirectement liés aux structures de données et aux interfaces de l'archivage numérique. La diffusion (recherche, découverte et visualisation d'objets archivistiques avec un instrument de recherche adéquat) et le *Contrôle de l'accès (Access Control)* (au sens de contrôle des autorisations d'accès) ne sont pas non plus ce qui nous occupe. Ces tâches sont aujourd'hui généralement accomplies par les systèmes d'information des Archives.

Le point de départ pour les spécifications est le *Paquet d'informations archivées (AIP)*, unité centrale des informations numériques archivées. En spécifiant leur AIP, les Archives concrétisent leurs exigences en matière d'archivage électronique. De cette spécification AIP découlent ensuite les autres interfaces: *Paquet d'informations à verser (SIP)*, *Information de Description* et *Paquet d'informations diffusé (DIP)*. Pour faciliter les références et la discussion, les différentes exigences seront numérotées dans le présent document.

Le CECO s'engage à continuer à tenir à jour sur la plateforme CECO.standards toutes les spécifications qui font partie de *bentō*<sup>4</sup>. Dans le futur proche, la compatibilité avec l'amont devrait pouvoir rester assurée, à condition que des extensions n'aient lieu que dans le domaine extra-obligatoire. La compatibilité amont ne pourra toutefois pas être assurée ad aeternam. On peut imaginer que sur de plus longues périodes, on en vienne à utiliser différentes spécifications d'AIP dans les mêmes Archives.

<sup>1</sup> CCSDS 650.0-B-1 (F): Modèle de référence pour un Système ouvert d'archivage d'information (OAIS). Livre bleu, Edition N°1, Mars 2005 (= ISO 14721:2003). [http://public.ccsds.org/publications/archive/650x0b1\(F\).pdf](http://public.ccsds.org/publications/archive/650x0b1(F).pdf)

<sup>2</sup> Exigences de base en matière d'archivage électronique. CECO 2009. [http://kost-ceco.ch/cms/index.php?minimal\\_specifications\\_fr](http://kost-ceco.ch/cms/index.php?minimal_specifications_fr).

<sup>3</sup> Voir [http://kost-ceco.ch/cms/index.php?project\\_arcun\\_fr](http://kost-ceco.ch/cms/index.php?project_arcun_fr).

<sup>4</sup> Voir [http://kost-ceco.ch/cms/index.php?standards\\_fr](http://kost-ceco.ch/cms/index.php?standards_fr). La gestion de la spécification SIP AFS incombe aux AFS.

## 2 AIP

Ce chapitre spécifie un framework pour l'élaboration / la définition d'un AIP. Dans ce contexte, le terme «framework» désigne un modèle d'ébauche (un «mode d'emploi de construction») d'un AIP. On pourra utiliser ce framework pour

- construire un AIP (comme de nombreuses archives ont entrepris de le faire dans de premières étapes d'archivage numérique, en produisant en quelque sorte des AIP à la main);
- implémenter un processus de création d'AIP (dans un système d'archivage numérique comme le DIR des AFS ou IBM-DIAS).

Le framework décrit un AIP sous une forme abstraite, il contient aussi certaines exigences que l'AIP doit remplir. En même temps, il laisse certaines directives ouvertes (au contraire d'une spécification formelle)<sup>5</sup>.

Le framework AIP est articulé en quatre domaines.

### 2.1 Généralités

1. Il s'agit de rechercher un AIP unifié pour tous les types et toutes les sources d'archives numériques.
2. Un nouvel AIP doit être généré en cas de migration de format de fichier ou de migration dans une nouvelle structure d'AIP. Le nouvel AIP contient une description de l'historique de la migration. Si l'AIP original est écarté après la migration, les données versées à l'origine doivent subsister dans le nouvel AIP. Le nouvel AIP doit continuer à vivre sous la cote existante; cela signifie que les cotes d'archives qui sont consignées dans l'AIP ne doivent pas être modifiées. L'ancien AIP doit contenir une cote dérivée de la cote d'origine. Il existe d'une part la possibilité de conserver toutes les versions d'un AIP. D'autre part, on peut envisager d'écarter les anciennes versions pour des raisons de place. Il est alors indispensable de conserver au minimum dans le nouvel AIP les fichiers versés à l'origine (dans l'intérêt d'un historique de migration), pour pouvoir encore corriger ultérieurement d'éventuelles erreurs de migration.

### 2.2 Fichiers primaires

1. Soit l'AIP ne comporte aucun fichier primaire, soit il en comporte un, soit encore il en comporte plusieurs<sup>6</sup>.
2. Par leur contenu, les fichiers primaires correspondent aux descriptions fournies dans le versement. Le nombre et les noms des fichiers concordent avec les informations correspondantes du versement.

<sup>5</sup> Une spécification concrète d'AIP pourrait faire l'objet d'un projet consécutif à *bentō*.

<sup>6</sup> Un AIP doit pouvoir ne contenir aucun ensemble de données primaires, mais uniquement des métadonnées, par exemple lorsque des documents analogiques ou hybrides sont pris en charge.

3. Les fichiers sont entiers (*integer*), c'est-à-dire que leurs sommes de contrôle concordent avec celles fournies par le service versant.
4. Les fichiers sont lisibles, c'est-à-dire qu'il n'existe aucune restriction d'accès, notamment aucun cryptage.
5. Les fichiers sont interprétables. En outre, ils sont enregistrés dans des formats acceptés par les Archives. Les formats de tous les fichiers sont identifiés, et mêmes validés par des contrôles par sondage<sup>7</sup>.
6. L'on peut renoncer à un contrôle antivirus. Un éventuel contrôle antivirus n'entraîne aucun changement des fichiers primaires.

### 2.3 Métadonnées

Dans l'AIP sont inventoriées des métadonnées de contenu, d'archivage, des métadonnées techniques, ainsi que des informations concernant des procédures, en provenance de diverses sources.

1. On y trouve au moins les métadonnées de contenu ci-dessous. Elles proviennent du service versant.
  - a) Liste des fichiers avec leurs noms et leurs sommes de contrôle.
  - b) Métadonnées obligatoires selon ISAD(G)<sup>8</sup>, à tous les niveaux de description (positions du système de classement): intitulé, date(s), importance matérielle et support (volume des fichiers et nombre de fichiers), nom du producteur, niveau (compte tenu de la réduction des redondances à l'intérieur de l'AIP).
  - c) Classification (catégorie, niveau de protection des données, statut de diffusion)
  - d) Plan d'archivage (au moins les chemins complets vers tous les dossiers contenus dans l'AIP)
  - e) Désignation des objets dans le système source (par ex. numéro de transaction, numéro de dossier, nom du fichier d'origine)
2. Le processus *Entrées* et toute migration ultérieure sont documentées dans l'AIP. En outre, au moins les informations suivantes concernant des procédures sont inventoriées:
  - a) Type d'action
  - b) Acteur
  - c) Date de l'action
  - d) (Seulement pour l'*Entrées*) Numéro de versement (c'est-à-dire une identification provisoire univo-

<sup>7</sup> Pour des raisons de productivité, on peut renoncer à valider entièrement tous les formats de fichiers.

<sup>8</sup> Brüning, Rainer, Heegewald, Werner u. Brübach, Nils (Übers. und Hrsg.), ISAD (G) – Internationale Grundsätze für die archivische Verzeichnung. Veröffentlichungen der Archivschule Marburg - Institut für Archivwissenschaft, Nr. 23. Marburg, 2002.  
[http://www.ica.org/sites/default/files/ISAD\\_2\\_DE.pdf](http://www.ica.org/sites/default/files/ISAD_2_DE.pdf).

que [GUID] de la livraison, qui caractérise la livraison [processus et fichiers correspondants] depuis le moment de sa planification jusqu'à l'archivage définitif de l'AIP)

- e) (Seulement pour l'*Entrées*) Service versant
- 3. Au titre de métadonnée archivistique, il faut au moins que la cote archivistique figure dans l'AIP.  
*Dans certaines archives, la cote est structurée par concaténation, c'est-à-dire qu'elle contient des composantes individuelle du classement des fonds. Ces composantes ne sont pas inventoriées ici à titre de documentation du classement, mais uniquement comme composantes de la cote.*
- 4. Il n'est pas obligatoire de conserver d'autres métadonnées techniques en dehors des sommes de contrôle des fichiers primaires que nous avons mentionnées. Si d'autres métadonnées techniques sont contenues dans l'AIP, elles sont enregistrées avec les métadonnées susmentionnées et sous la même forme.
- 5. Toutes les métadonnées obligatoirement inventoriées dans l'AIP sont consignées sous une forme documentée et interprétable par des machines. La documentation correspondante doit être enregistrée dans l'AIP lui-même.
- 6. Les métadonnées reflètent la date du versement, ou plus exactement le moment où le traitement et l'utilisation des documents ont pris fin au lieu de provenance. C'est pourquoi il n'est pas admissible de procéder ultérieurement dans l'AIP à une modification dépendant de l'instrument de recherche.

## 2.4 Conteneur

- 1. Au minimum, l'AIP est organisé comme un conteneur logique, c'est-à-dire comme une liaison entre données primaires et données correspondantes qui permet à

## 3 SIP

Ce chapitre spécifie une interface de versement vers les archives numériques.

### 3.1 Situation initiale

L'OASIS définit l'*Entrées* comme l'unité fonctionnelle qui reçoit un SIP du producteur de données et qui prépare les contenus de ce SIP pour un enregistrement et une gestion au sein des Archives. Le processus *Entrées* génère notamment un AIP.

Ce processus est très superficiel dans la plupart des implémentations de l'OASIS que l'on connaît dans les archives et les institutions similaires. Cela tient au fait que le SIP est repris en tant qu'AIP, pratiquement sans changement. Le «processus de transformation» d'un SIP en AIP se limite donc à en vérifier l'intégrité et à l'enrichir de quelques métadonnées.

Les étapes plus épineuses et plus décisives du passage des données d'origine à l'AIP sont alors avancées au stade de la constitution du SIP. C'est pourquoi la no-

tout moment de restaurer l'archive indépendamment, sans informations propriétaires externes. La structure interne de ce conteneur prévoit un secteur pour les données primaires et un secteur pour les métadonnées. Si en outre des fichiers destinés à la documentation des données primaires sont transmis, ces fichiers sont enregistrés avec les données primaires. Si une documentation de l'AIP est disponible, elle doit être enregistrée avec les métadonnées. Le nom attribué à la structure interne doit soit être auto-explicatif, soit découler d'autres normes (référéncées).

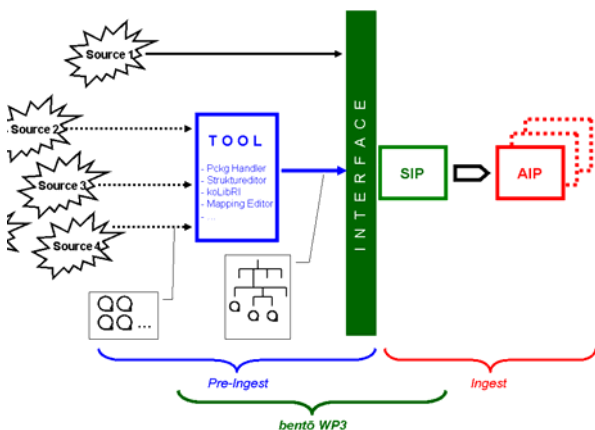
- 2. Les métadonnées sont liées de manière univoque avec les données primaires auxquelles elles se réfèrent.
- 3. Dans la mesure du possible, on conservera les structururations internes des données primaires. Au besoin, les noms de fichiers et de répertoires seront normalisés pour qu'il soit possible de respecter les restrictions techniques du format de conteneur choisi. (Concrètement, par exemple, les noms de chemin et de fichier ne doivent contenir que des signes admis sous tous les systèmes d'exploitation usuels.) En cas de normalisation, les noms de fichiers et de répertoires sont documentés dans l'AIP.
- 4. Il existe dans le conteneur un point d'introduction défini. Il s'agit d'un fichier ayant un nom univoque, à un emplacement fixe, et qui renvoie à des informations guides.
- 5. L'AIP est désigné par un GUID. Les Archives garantissent la mise en équivalence des cotes d'archives avec le GUID de l'AIP.
- 6. Un mécanisme permettant de garantir l'intégrité du conteneur (intégrité des trains de bits) est implémenté.

tion de *pré-Entrées* a fait son apparition. En général, on utilise pour cela un outil<sup>9</sup> qui constitue, à partir de données diversement structurées et provenant de diverses sources, un SIP défini en fonction d'une interface (ou encore, le système de provenance des documents est équipé d'une interface pour la création directe d'un SIP). Cet outil peut même servir à éditer dans les archives le SIP ainsi constitué, et également à naviguer dans l'AIP, qui lui est identique pour l'essentiel, et cela sans devoir recourir à un éditeur XML<sup>10</sup>.

<sup>9</sup> AFS: Package Handler; kopal/DIAS: koLibRI; BArch: Struktureditor; RODA: Mapping Editor.

<sup>10</sup> Cette possibilité d'explorer confortablement le SIP et l'AIP et de pouvoir les éditer nous paraît trop déterminante pour que l'on puisse se passer d'un tel outil de structuration, même si l'implémentation directe d'une in-

Ces observations montrent bien que les spécifications de *bentō* ne peuvent pas se limiter à la transformation du SIP en AIP. Elles doivent aussi comprendre le processus de *pré-Entrées*. Le but visé est de définir une interface SIP qui, d'une part, puisse être maniée par les différents outils (et directement à partir d'une application GEVER), et d'autre part soit suffisamment détaillée pour qu'il soit possible de former, à partir d'un SIP ainsi préparé, un AIP correspondant au framework pour AIP *bentō* (voir chapitre 2). Le SIP doit pouvoir être transformé en un AIP sans perte de données ni incohérences; dans cette mesure, le SIP doit posséder une structure fondamentalement similaire à celle d'un AIP, mais il peut s'en différencier par sa granularité.



Processus *pré-Entrées* et *Entrées*

## 3.2 Analyse du SIP des AFS

Nous utiliserons comme base pour l'interface SIP *bentō* la spécification de l'interface SIP des AFS (*Submission Information Package v1.0* de septembre 2009<sup>11</sup>). Dans le cadre de *bentō*, l'on a examiné si toutes les exigences du framework AIP spécifiées au chapitre 2 étaient satisfaites, c'est-à-dire si un SIP ainsi spécifié pouvait être transformé en un AIP sans perte de données ni incohérences<sup>12</sup>. Il s'avère que quelques adjonctions suffisent pour que la spécification SIP AFS satisfasse aux exigences du framework pour AIP *bentō*. Les résultats détaillés de cette analyse ainsi que les adjonctions nécessaires sont exposés ci-après.

### 3.2.1 Généralités

Les exigences AIP sont énumérées et référencées une par une. Il faut à ce sujet savoir ce qui suit:

- La spécification SIP des AFS comprend le document de spécification proprement dit, le *Diction-*

terface SIP dans le système source paraît rendre la fonction structurante de cet outil superflue.

<sup>11</sup> Voir <http://www.bar.admin.ch/dienstleistungen/00895/00897/>, citée ci-après sous la dénomination *SIP AFS*.

<sup>12</sup> Ceci également au cas où la granularité changerait lors de la transition entre le SIP et l'AIP.

*naire de données (Data Dictionary)*, avec la documentation correspondante ainsi que les schémas XML. C'est en premier lieu la spécification qui est déterminante; ceci tout particulièrement lorsque des exigences indispensables de la spécification ne sont pas représentées dans des *mandatory elements* des schémas XML. C'est pourquoi nous renverrons ci-après toujours aux identificateurs des exigences respectives de la spécification.

- Certaines exigences du framework AIP ne concernent que le processus *Entrées* ou l'AIP lui-même, mais pas le SIP. Ces exigences sont sommairement récapitulées à la fin de l'analyse.
- La spécification SIP *bentō* est une extension de la spécification SIP AFS, à raison d'une adjonction (*bentō* 3.2.3.1.f). Une rétrocompatibilité du SIP *bentō* vers le SIP AFS est réalisée: un SIP AFS valide valide également par rapport à la spécification SIP *bentō* (mais l'inverse n'est pas forcément vrai).

### 3.2.2 Fichiers primaires

1. (*bentō* 2.2.1) Les fichiers primaires sont placés dans le répertoire «content» du SIP AFS (SIP AFS S\_5.4-6). La spécification n'exclut pas des SIP vides.
2. (*bentō* 2.2.2) Dans le SIP AFS, le nombre de fichiers (implicitement) et leurs noms sont transmis (SIP AFS M\_4.7-1).
3. (*bentō* 2.2.3) Dans le SIP AFS, des algorithmes de vérification et des sommes de contrôle doivent être entrés en mémoire pour les différents fichiers (SIP AFS M\_4.11-1)<sup>13</sup>.

La validation de l'intégrité des fichiers fait partie du processus *Entrées*.

4. (*bentō* 2.2.4) Dans le SIP AFS, les fichiers ne doivent être ni cryptés, ni protégés par mot de passe (SIP AFS A\_2.6-1 et A\_2.6-2).

La validation de la lisibilité fait partie du processus *Entrées* et est indépendante du SIP.

5. (*bentō* 2.2.5) Dans le SIP AFS, les données primaires doivent être intégrées dans un format adapté à l'archivage et approprié au document, conformément aux spécifications de la fiche d'information «Formats de fichier adaptés à l'archivage, Normes et standards pour l'archivage de documents numériques» des AFS (SIP AFS P\_3.2-1).

L'identification et la validation des formats de fichiers font partie du processus *Entrées* et sont indépendants du SIP.

### 3.2.3 Métadonnées

1. (*bentō* 2.3.1) Métadonnées de contenu.

<sup>13</sup> Dans le schéma XML actuel, les éléments <algorithme de contrôle> et <somme de contrôle> ne sont pas *mandatory*. Les AFS sont conscientes de cette incohérence, qui sera corrigée dans la prochaine version du schéma.

- a) *Liste des fichiers avec leurs noms et leurs sommes de contrôle*: dans le SIP AFS, tous les répertoires et les fichiers du paquet doivent être cités dans la table des matières du fichier `meta-data.xml`, avec le nom du fichier original et la somme de contrôle (SIP AFS M\_4.7-1).
- b) *Métadonnées obligatoires selon ISAD(G) à tous les niveaux de description (à toutes les positions du système de classement): titre, date(s), importance matérielle (volume du fichier et nombre de fichiers), provenance, niveau*: dans le SIP AFS, la provenance est obligatoirement définie au niveau du versement (et sera héritée conformément à ISAD[G]); le titre est consigné au niveau dossier et document, les dates au niveau du dossier; le volume de fichier et le nombre de fichiers sont consignés implicitement; le niveau est représenté dans le SIP par la structure du système de classement (voir *Dictionnaire de données*).
- c) *Classification (catégorie, niveau de protection des données, statut de diffusion)*: les indications concernant les délais de protection des documents versés doivent être consignées dans les métadonnées du SIP. En provenance du système GEVER, on reprend en outre les métadonnées concernant la protection des données, la classification et le statut de diffusion (SIP AFS M\_4.9-1). (Dans le schéma XML, ces éléments ne sont pas *mandatory*.)  
Nous recommandons de sélectionner les valeurs de l'élément `<schutzfristenkategorie>` à partir d'une liste contrôlée de valeurs qui dépend de la législation applicable en matière d'archives.
- d) *Plan d'archivage (au moins les chemins complets vers tous les dossiers contenus dans le versement)*: s'il s'agit d'un SIP GEVER, tout le système de classement (avec toutes les positions du système) doit être repris dans le SIP (SIP AFS M\_4.8-2).
- e) *Désignation des objets dans le système source (numéro de transaction / numéro de dossier / référence du dossier ainsi que nom d'origine du fichier)*: dans le SIP AFS, la référence au dossier est obligatoire (sous forme de numéro de transaction ou de dossier). Dans le schéma XML, le nom d'origine du fichier n'est pas obligatoire parce qu'il n'est inventorié que s'il diffère du nom actuel (voir *Dictionnaire de données*) (SIP AFS S\_5.3-5).
- f) *En plus de la spécification actuelle des AFS, le SIP doit fournir une possibilité de consigner d'autres métadonnées de n'importe quel type en provenance du système source sous forme de couples clé-valeur*. L'on risque d'ouvrir ainsi la voie à un développement incontrôlé de métadonnées dans le SIP (et en conséquence dans l'AIP). Mais la possibilité de consigner dans le SIP et dans l'AIP d'autres métadonnées que le mini-

num spécifié répond à un besoin largement exprimé et est donc incontournable.

2. (*bentō 2.3.5*) Les métadonnées du SIP AFS se trouvent toutes dans le fichier `meta-data.xml`, conformément aux schémas XML (SIP AFS M\_4.1-1).

### 3.2.4 Conteneur

1. (*bentō 2.4.1*) Cette exigence ne concerne que l'AIP. Le principe est le suivant: puisque le SIP AFS, en tant que paquet logique, est défini avec un répertoire *header* pour les métadonnées et un répertoire *content* pour les données primaires, cette exigence a déjà été satisfaite en cas de transmission sans changements du SIP sous forme d'AIP. (Il faut cependant tenir compte du fait que la subdivision ici exigée entre données primaires et métadonnées ne correspond pas dans tous les cas à la subdivision entre *header* et *content*, qui tire son origine de l'informatique.)
2. (*bentō 2.4.2*) Dans le SIP AFS, les fichiers primaires sont référencés au moyen d'un élément `<DateiRef>` dans les métadonnées (non *mandatory*) (SIP AFS M\_4.12-1).
3. (*bentō 2.4.3*) Dans le SIP AFS, les noms de fichier et de répertoire sont normalisés selon des critères précis (SIP AFS, Appendix C). Les critères appliqués garantissent que les noms de fichier et de répertoire correspondent aux restrictions techniques de tous les systèmes les plus répandus.
4. (*bentō 2.4.4*) La structure interne du SIP AFS est définie de manière fixe (SIP AFS S\_5.4-1, S\_5.4-2, S\_5.4-3, S\_5.4-4, S\_5.4-5, S\_5.4-6, S\_5.4-7). Le point d'introduction est le fichier `meta-data.xml`, dans le répertoire *header* (SIP AFS S\_5.4-4).

### 3.2.5 Exigences hors de propos pour le SIP

1. Les exigences *bentō 2.2.6, 2.3.2 et 2.3.3* concernent le processus *Entrées* et l'AIP indépendamment du SIP.
2. Les exigences *bentō 2.3.4, 2.4.5 et 2.4.6* concernent uniquement l'AIP indépendamment du SIP.

### 3.2.6 Remarques explicatives complémentaires

1. Pour les métadonnées concernant des processus, des procédures etc. tirées du système de provenance, il existe au niveau dossier du SIP AFS la métadonnée `<Dossier_Vorgang>`, dans laquelle ces informations peuvent être reprises dans un champ de texte libre. Une autre possibilité envisageable consiste à garantir les procédures du système de provenance sous forme de donnée primaire dans le SIP. Quant aux procédures qui concernent respectivement le SIP et l'AIP (procédures dites archivistiques), les AFS les consignent dans des métadonnées spécialement prévues à cet effet.
2. Dans le SIP AFS, un versement ne peut venir que d'une provenance unique.

3. Les cotes numériques sont actuellement ignorées par les AFS. Si des documents signés numériquement sont transmis aux Archives, il devrait être possible de documenter la cote perdue dans des champs de métadonnées supplémentaires (le cas échéant sous forme de couples clé-valeur).

### 3.3 Réflexions sur la granularité des SIP et des AIP

Les Archives représentées dans le projet *bentō* souhaitent des granularités différentes pour leurs SIP et leurs AIP. Ainsi, les AFS ont choisi la granularité versement pour les SIP et les AIP. D'autres Archives projettent d'enregistrer des AIP de granularité dossier. Une granularité plus fine a également été envisagée pour le versement, c'est-à-dire pour le SIP.

Les considérations issues de la science de l'archivage n'ont pas débouché sur une granularité standard pour l'AIP ni pour le SIP. L'on se rend compte, en revanche, que des considérations pratiques jouent un rôle essentiel.

Les conclusions suivantes sont consensuelles:

## 4 Information de Description

Dans la terminologie de l'OAIS, l'*Information de Description* est une interface qui relie le processus *Entrées* et le SIP au *Gestion de données*, c'est-à-dire à le *Système d'archivage d'information (AIS)* et à l'instrument de recherche. Cette interface doit servir d'intermédiaire entre deux ensembles de données différents:

- les métadonnées descriptives, qui font partie du SIP / de l'AIP et proviennent du système versant (voir à ce sujet la liste des métadonnées de l'AIP)
- l'inventaire archivistique, basé sur la norme internationale ISAD(G), laquelle est implémentée dans les différents AIS.

### 4.1 Interfaces AIS existantes

Les AIS et les instruments de recherche utilisés par les Archives membres du CECO<sup>14</sup> offrent parfois déjà des interfaces spécifiées pour l'*Information de Description*. Il existe en outre pour certains AIS des modules de transmission qui supportent la transmission automatisée de métadonnées en provenance du producteur de données. Un aperçu de ces interfaces et de ces modules a montré ce qui suit:

- Divers AIS offrent un *Mapping Editor* qui sert d'intermédiaire entre les métadonnées provenant des systèmes actifs et les métadonnées de description selon ISAD(G).

<sup>14</sup> scopeArchiv, CMISTAR, Augias, Faust, BASIS+, Flora Archives.

1. La granularité du SIP et de l'AIP doit être définie en fonction de la charge de travail et des coûts du traitement.
2. La spécification SIP doit autoriser toutes les granularités, depuis celle du dossier jusqu'à celle du versement, et supporter la migration granulaire, c'est-à-dire la possibilité de modifier la granularité. Cette modification peut avoir lieu au moment de la transition entre SIP et AIP ou entre AIP et DIP, mais aussi entre différentes versions d'un SIP (dans le cadre de son traitement aux Archives). Elle peut s'effectuer dans deux sens: vers un grand paquet par fusion de plusieurs petits paquets, ou vers plusieurs petits paquets par fractionnement d'un grand paquet. L'enregistrement, tel que l'exige le framework AIP, du chemin entier vers chaque dossier du SIP/AIP est une condition de cette procédure.

### 3.4 Développements

Les processus et les outils pour générer et valider des SIP n'entraient pas dans le cadre du projet *bentō*. La nécessité de tels instruments est cependant expressément soulignée. Leur élaboration devrait être visée dans le cadre d'un projet ultérieur.

- Ce *mapping* (mise en équivalence) est en général possible sur un seul niveau hiérarchique à la fois; la transmission de métadonnées structurées hiérarchiquement nécessite généralement un *workaround* (solution de contournement).
- Plusieurs AIS peuvent prendre en charge des métadonnées en format XML et enregistrer un *mapping* correspondant.
- Les métadonnées en format XML devraient en principe permettre une automatisation de la transmission. Beaucoup de travail manuel reste cependant encore nécessaire actuellement.

Les fichiers de transmission XML spécifiques aux produits doivent être remplacés par un format normalisé. Le résultat visé est une automatisation aussi complète que possible du processus.

### 4.2 Contenu de l'interface pour l'Information de Description

1. L'interface pour les informations descriptives comprend onze champs obligatoires de métadonnées, qui appartiennent à trois catégories différentes:
  - a) les champs obligatoires ISAD(G);
  - b) d'autres champs ISAD(G) qui doivent être déclarés obligatoires pour les documents numériques (et être disponibles dans l'AIP, c'est-à-dire pouvoir être transmis automatiquement);
  - c) d'autres métadonnées en dehors d'ISAD(G), qui sont nécessaires pour les documents numériques.



Ces champs de métadonnées sont décrits ci-dessous. Des raisons et des précisions sont également mentionnées. Là où ce n'est pas le cas, nous renvoyons à la définition correspondante dans ISAD(G) ainsi qu'au *Dictionnaire de données* annexé au présent document.

#### Champs obligatoires selon ISAD(G)

- 1.1 Référence
- 1.2 Intitulé / analyse
- 1.3 Date(s) de création / durée  
*Cet élément comporte trois sous-éléments: «exactement», «depuis» et «jusqu'à», toujours dans la classe «année» ou «date». Pour chaque document, une date doit être reprise dans le SIP (si possible la date de clôture). Aux niveaux de description les plus élevés, l'élément «Date(s) de création / durée» est concaténé à partir des données des niveaux de description ou des documents hiérarchiquement subordonnés.*
- 1.4 Niveau de description
- 1.5 Importance matérielle et support de l'unité de description (quantité, volume ou dimension)  
*Cet élément doit désormais être spécifié pour les documents numériques. Les unités de mesure qui entrent en ligne de compte sont la taille totale du fichier (ou la taille de fichier de l'AIP) ou le nombre de fichiers.*
- 2.1 Nom du producteur

#### Autres champs ISAD(G) déclarés obligatoires

- 2.4 Modalités d'entrée  
*Cet élément est considéré comme nécessaire car il faut partir du fait que, dans le monde numérique plus souvent que dans le monde analogique, le service versant n'est pas identique à la provenance.*
- 4.1 Conditions d'accès  
*Il est possible de reprendre dans le SIP différents facteurs qui ont une influence sur les conditions d'accès: confidentialité (ou catégorie de classification), niveau de protection des données (contient ou non des données personnelles dignes de protection), statut de diffusion (le document ou la rubrique a été désigné comme accessible au public selon la Loi sur la transparence), catégorie de délai de protection (référence à un article de loi), délai de protection (délai concret en années). Ces informations ne dé-*

*coulent pas automatiquement des données primaires et doivent donc être transmises dans l'interface DI. La détermination concrète des dispositions d'accès dépend de diverses prescriptions spécifiques aux Archives respectives.*

- 4.4 Caractéristiques matérielles et contraintes techniques  
*Cet élément doit être utilisé pour identifier les documents comme numériques, analogiques ou hybrides.*

#### Métadonnées supplémentaires tirées de l'AIP

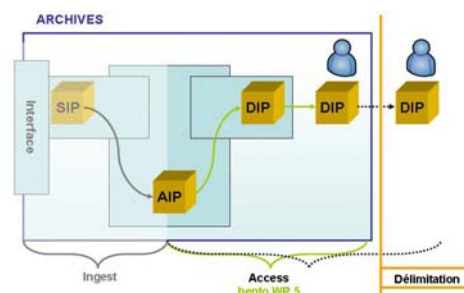
- Numéro de référence dans le système original  
*Cet élément est nécessaire pour assurer une identification rapide des documents archivés que l'organisme de provenance demande à consulter.*
  - Référence à l'AIP
2. En outre, l'interface comprend à titre facultatif tous les autres champs ISAD(G).
  3. L'idée de reprendre également en bloc dans l'AIS toutes les métadonnées GEVER enregistrées en tant que couples clé-valeur (soit en spécifiant un formulaire descriptif pour chaque versement, soit par une liste en texte libre dans un champ commentaire) a été rejetée. Ces métadonnées ne sont pas des métadonnées descriptives au sens archivistique, aussi n'ont-elles pas leur place dans un système d'information d'archives (mais elles restent bien sûr inventoriées dans l'AIP et peuvent être diffusées dans le DIP). Au sujet de l'exception faite pour les métadonnées spécifiques au format, voir *bentō* 4.2.4.
  4. Comme avec EAD ou METS, il doit être possible d'étendre le schéma pour décrire d'autres métadonnées spécifiques au format, selon la catégorie de format. Ces métadonnées doivent être transmises sous forme de couples clé-valeur. Des groupes de travail de relais pourront élaborer des définitions correspondantes.

### 4.3 Dictionnaire de données et schéma XML

L'interface que nous avons résumée ici est spécifiée en détails dans un *Dictionnaire de données (Data dictionary)* ainsi que par un schéma XML. Ces documents sont annexés à la présente spécification.

## 5 DIP

On peut définir comme une transformation la rétro-conversion de l'AIP dans un DIP, avec un éventuel changement de granularité selon le point de mire de l'utilisateur.



### *Définition du DIP*

1. Un DIP est un paquet élaboré temporairement sur la base d'un AIP pour la diffusion d'archives. C'est-à-dire qu'un DIP est en principe créé quand on a besoin du contenu. Les DIP dont le contenu est souvent requis peuvent aussi être conservés dans une mémoire ad hoc. Ils servent toutefois uniquement à la diffusion et ne sont soumis à aucune mesure de pérennisation.
2. Un DIP contient l'archive numérique que l'utilisateur a commandée; il contient à la fois les données primaires et les métadonnées correspondantes, dans un même paquet.
3. Un DIP et les documents qu'il contient (données primaires et métadonnées) sont élaborés pour être «humainement lisibles»; c'est-à-dire qu'un utilisateur peut comprendre sans outils spéciaux les documents qui sont mis à sa disposition (voir toutefois 5.15).
4. Un DIP peut correspondre à l'unité que l'on peut commander dans le *repository* des archives (par ex. dossier). Il est toutefois également possible d'empaqueter dans un DIP plusieurs dossiers ou des collections complètes de documents, par exemple des séries provenant de différents versements (DIP sur mesure).
5. Un DIP est toujours une copie des documents de l'AIP. Un utilisateur des archives n'a jamais directement accès à un AIP.
6. Pour les DIP qui contiennent des documents de différents versements / AIP, une exigence de base est que ces AIP aient au stade pré-archivistique la même structure, ou tout au moins une structure compatible, qui puisse être également versée dans le DIP.

### *Transformation AIP → DIP*

7. La transformation d'un AIP en un DIP est assurée par les Archives. Les préliminaires à la mise en œuvre technique de cette transformation sont une tâche des Archives.
8. Il existe plusieurs variantes pour opérer la transformation de l'AIP en DIP. Les Archives peuvent choisir en fonction de leurs exigences:
  - a) La transformation de l'AIP en DIP se fait en mode 1/1. Les documents du DIP correspondent alors exactement à ceux qui sont dans l'AIP.
  - b) Seule une partie des documents de l'AIP sont repris dans le DIP. Mais tous les documents du DIP proviennent du même AIP.
  - c) Des documents provenant de différents AIP (complets ou partiels) sont rassemblés dans un unique DIP. Cela n'est possible que si les métadonnées du système de classement sont identiques dans les AIP concernés.
9. La granularité de l'AIP a une influence déterminante sur l'implémentation du DIP, c'est-à-dire que les possibilités de constituer un DIP dépendent aussi de la structure, de la construction et des contenus de l'AIP.

10. La transformation de l'AIP en DIP suit les avancées des possibilités de diffusion d'archives numériques. Elle peut changer au fur et à mesure des nouvelles exigences et possibilités (des archives et des utilisateurs).

### *Données primaires du DIP*

11. Le DIP peut contenir les données primaires dans un autre format que dans l'AIP. Exemples:
  - a) Au moment de la constitution du DIP, les données primaires peuvent être converties en un format convivial.
  - b) Les données primaires peuvent être pourvues d'un filigrane.

### *Métadonnées du DIP*

12. Une grande partie des métadonnées du DIP provient de l'AIP.
  - a) Métadonnées descriptives et techniques associées aux données primaires;
  - b) Métadonnées associées au système de classement de la source, si elles sont nécessaires à la compréhension des données contenues dans le DIP.
13. En plus des données, primaires et métadonnées, provenant de l'AIP, le DIP contient les métadonnées suivantes, qui renseignent avant tout sur la provenance et la production du DIP et sur la possibilité de reconstituer la transformation AIP → DIP:
  - a) Date de génération: informations sur le moment auquel le DIP a été généré; date/horodatage (obligatoire)
  - b) Préparateur: qui (aux Archives) a constitué le DIP? (facultatif)
  - c) Producteur du DIP: informations sur la provenance du DIP: dans quelles archives se trouvaient les documents? (obligatoire)
  - d) Cote(s) des documents commandés contenus dans le DIP: sert à garantir la possibilité de citation et la possibilité d'attester pour les utilisateurs que les documents commandés ont été livrés. (obligatoire)
  - e) Auteur de la commande d'un DIP: pour qui le DIP a-t-il été constitué? (facultatif)

14. Il n'est pas obligatoire que le DIP contienne des métadonnées de l'instrument de recherche / du système d'information des archives; c'est-à-dire que la classification des fonds et les métadonnées qu'elle contient ne sont généralement pas reproduites dans le DIP parce qu'elles servent uniquement à la diffusion, et non à l'utilisateur de l'archive.

### *Forme du DIP*

15. Il existe différentes possibilités de préparation d'un DIP. Elles dépendent fortement des exigences des archives et des exigences des utilisateurs respectifs:
  - a) Un DIP est une unité que les utilisateurs peuvent emporter avec eux si les conditions d'utilisation de l'archive qu'il contient le permettent.



- b) Un DIP est au minimum un conteneur (par ex. un ZIP).
- c) Dans la mesure du possible, un DIP doit être autonome. L'initiation au DIP et l'accès aux documents doivent être possibles sans information extérieure au DIP. Les possibilités suivantes de soutien à l'utilisation (aide introductive) sont envisageables:
  - o Un outil de visualisation actuel est fourni avec le DIP.
  - o Il existe une «page d'entrée» (par ex. en format HTML). Le DIP est préparé de manière à pouvoir être parcouru avec un navigateur usuel. En alternative, un fichier «autostart» peut être fourni pour le DIP.
  - o Une aide minimale éventuelle, mais à éviter si possible, serait un fichier `readme.txt` qui expliquerait la manière de procéder avec le DIP.
- d) Il peut y avoir différentes «possibilités d'aperçu» des contenus du DIP: par ex. des onglets pour les images.

*Remarques au sujet de la notion de convivialité: Pour les documents texte et image, il est plus facile d'implémenter un accès utilisateur auto-explicatif. Pour les données plus complexes, par exemples celles provenant de bases de données ou d'applications spécialisées, les modes d'emploi doivent déjà être préparés par les services versants au stade de la constitution du SIP. En outre, de telles données peuvent nécessiter de mettre des connaissances de praticien à la disposition des utilisateurs du DIP.*

*Scénarios de mise à disposition*

- 16. La politique de préparation des Archives a une influence sur l'accessibilité des documents du DIP pour les utilisateurs. Mais cela n'influe que faiblement sur la forme du DIP. Chaque DIP en soi devrait être conçu pour pouvoir être livré en tant qu'objet.
  - a) Avec un DIP, les documents peuvent être téléchargés et emportés par l'utilisateur.

- b) L'utilisateur peut uniquement visionner les documents et les métadonnées correspondantes, mais ni les télécharger ni les réaffecter ailleurs (hors du DIP).

*Contrôle d'accès*

- 17. Le contrôle de l'accès aux documents numériques doit être implémenté par les Archives indépendamment du DIP. Le DIP est uniquement une forme d'utilisation; il ne «sait» rien des droits d'accès et d'utilisation. Ceux-ci doivent être fixés du point de vue organisationnel et implémentés et gérés techniquement dans d'autres systèmes (par ex. AIS). Cependant, comme dans un SIP, les conditions d'accès peuvent être transmises avec le DIP, dans les métadonnées.
- 18. Les Archives doivent adopter des réglementations pour définir dans quels cas le contrôle d'accès peut passer des Archives à l'utilisateur. Ces réglementations dépendent des prescriptions des Archives en ce qui concerne la remise de copies de documents à des utilisateurs (par ex. imprimer, emporter sous forme numérique) et les droits d'accès de ces derniers.
- 19. Nous recommandons aux Archives d'intégrer dans le DIP un *disclaimer* (décharge de responsabilité) concernant l'utilisation des archives du DIP. Ce *disclaimer* devrait encore une fois rendre l'utilisateur attentif aux conditions cadres de l'utilisation des documents du DIP. On garantit de cette manière que l'utilisateur a toujours accès aux informations lui disant ce qu'il a le droit de faire avec les contenus du DIP et ce qui n'est pas autorisé.

*Authenticité des documents délivrés (DIP)*

- 20. Dans leurs processus de constitution, les Archives garantissent à l'utilisateur d'un DIP que les documents mis à disposition sont bien les documents authentiques qui ont été commandés.

## 6 Annexes

Annexe 1:

*Dictionnaire de données pour l'Information de Description*

Annexe 2:

*Définition du schéma XML pour l'Information de Description*

Annexe 3:

*Trajet bentō-SIP → Information de Description*