



Générateur de SIP - Mode d'emploi

Version du 12/12/2016

Valeurs Immatérielles Transférées aux Archives pour Mémoire

Suivi des versions du document

Version	Date	Auteur	Modifications
V0.1	12/12/2016	EV	Création du document
V0.2	13/01/2017	MR	Mise à jour licence

Licence

- Ce document est distribué sous les termes de la licence Creative Commons Attribution - Partage dans les Mêmes Conditions 3.0 France (CC BY-SA 3.0 FR).

Sommaire de la présentation

1/ Présentation du générateur

2/ Fonctionnement simple du générateur

3/ Fonctionnement avancé du générateur

Annexe : Rappels sur le SEDA 2.0.

Présentation du générateur

Objectifs de l'outil

- Faciliter la génération des jeux de tests pour les membres de l'équipe Vitam, les ministères porteurs et les partenaires
- Générer automatiquement à partir d'une arborescence de fichiers des SIP pouvant être pris en charge par la solution logicielle Vitam :
 - Compatible NF Z44-022 et standard SEDA v2.0 (conforme schéma .xsd du SEDA 2.0.)
 - Conforme au document de spécification des SIP spécifique à la solution logicielle Vitam
 - Sans nécessairement avoir besoin d'utiliser un éditeur xml pour créer le bordereau
 - Permettant de générer rapidement un SIP avec des milliers d'unités d'archives et des milliers de fichiers
 - De manière fiable (ne pas copier à la main l'empreinte)

Alimentation du bordereau (1)

- Pour les fichiers
 - Calcul de l'empreinte (avec l'algorithme de hachage paramétrable, par défaut SHA-512) et écriture dans le bordereau
 - Calcul de la taille du fichier et écriture dans le bordereau
 - Récupération dans le bordereau du nom d'origine du fichier et de sa date de dernière modification (FileInfo)
 - Définition d'un usage par défaut pour les fichiers (original numérique = BinaryMaster)
 - Identification du format du fichier en utilisant l'outil Siegfried

Alimentation du bordereau (2)

- Pour l'arborescence d'unités de description
 - Création d'une arborescence d'unités archivistiques reprenant l'arborescence du système de fichiers
 - Création des liens entre unités archivistiques et fichiers numériques
 - Indication de niveaux de description par défaut : RecordGrp pour les répertoires, Item pour les fichiers
 - Alimentation automatique du bordereau avec les informations suivantes récupérées de l'arborescence de fichiers
 - *Dates (date de modification des fichiers, dates extrêmes des répertoires)*
 - *Titre (répertoire, fichier)*

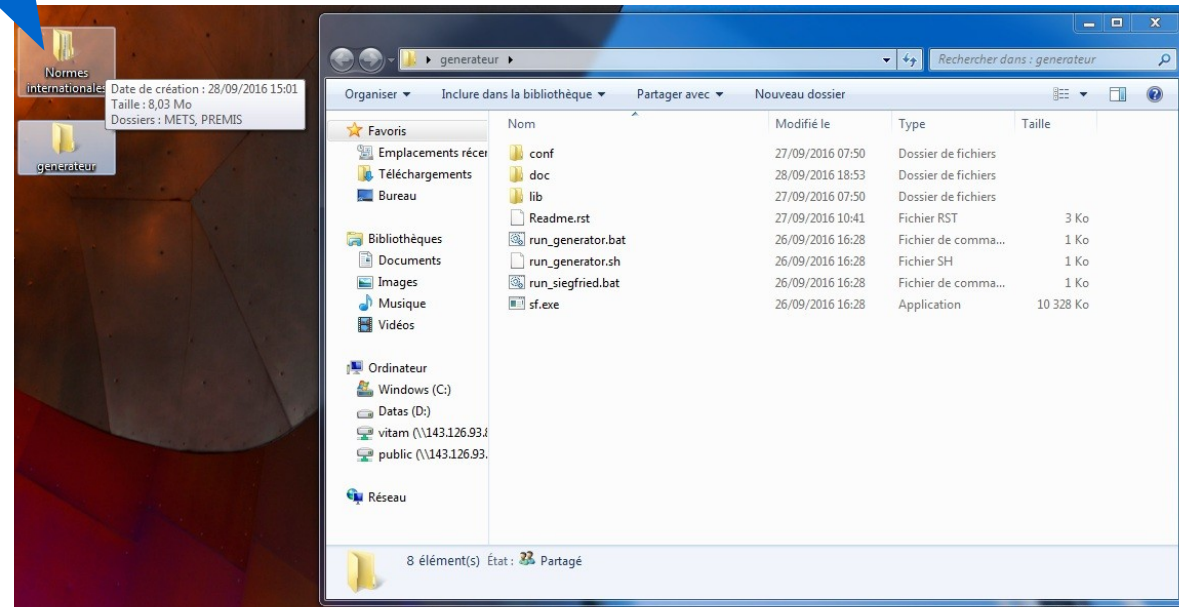
Paramétrage de l'alimentation du SIP

- Paramétrage de l'en-tête du bordereau
 - Alimentation de certains champs : Identifiant du SIP, commentaire, contrat d'entrée, identifiant du service versant, identifiant du service d'archives, listes de codes
 - Définition des fichiers que l'utilisateur ne souhaite pas mettre dans le SIP : par exemple Thumbs.db, *.vcf
- Paramétrage des métadonnées des répertoires
 - Alimentation possible de tous les champs existants dans l'ontologie définie dans le SEDA 2.0.
- Autres paramétrages possibles :
 - Rejet des fichiers non reconnus par l'outil d'identification des formats Siegfried. Par défaut, ces fichiers sont inclus dans le SIP
 - Rejet des fichiers ayant une extension contenant un caractère « URL-encoded » de type « + ». Par défaut, ces fichiers sont inclus dans le SIP
 - Choix d'un autre algorithme de calcul d'empreinte des fichiers que l'algorithme par défaut (SHA-512)

Fonctionnement simple du générateur

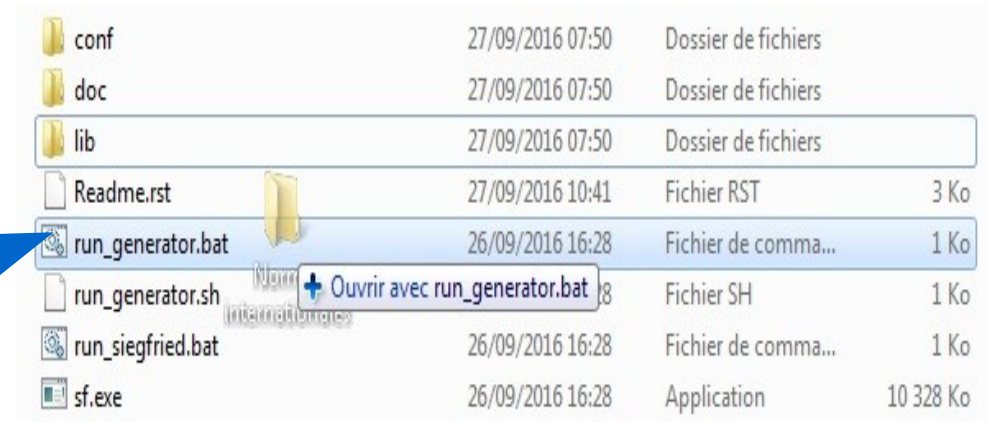
Étape 1 - Identifier le dossier à archiver

- Choisir dans le navigateur la racine de l'arborescence à intégrer dans le SIP
- Attention : si un seul fichier doit être archivé, il convient de l'enregistrer dans un répertoire

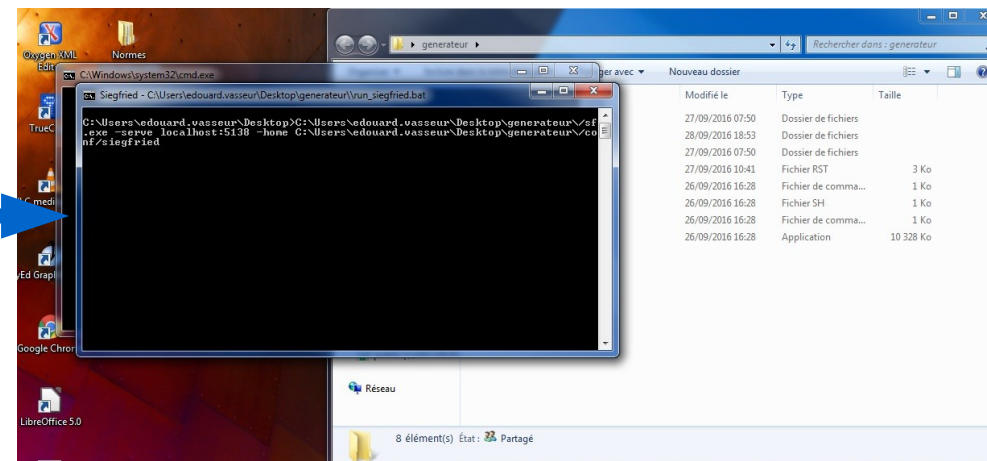


Étape 2 - Lancer l'opération de création du SIP

- Ouvrir le répertoire contenant le générateur
- Faire glisser le dossier sélectionné vers le fichier « run_generator.bat »



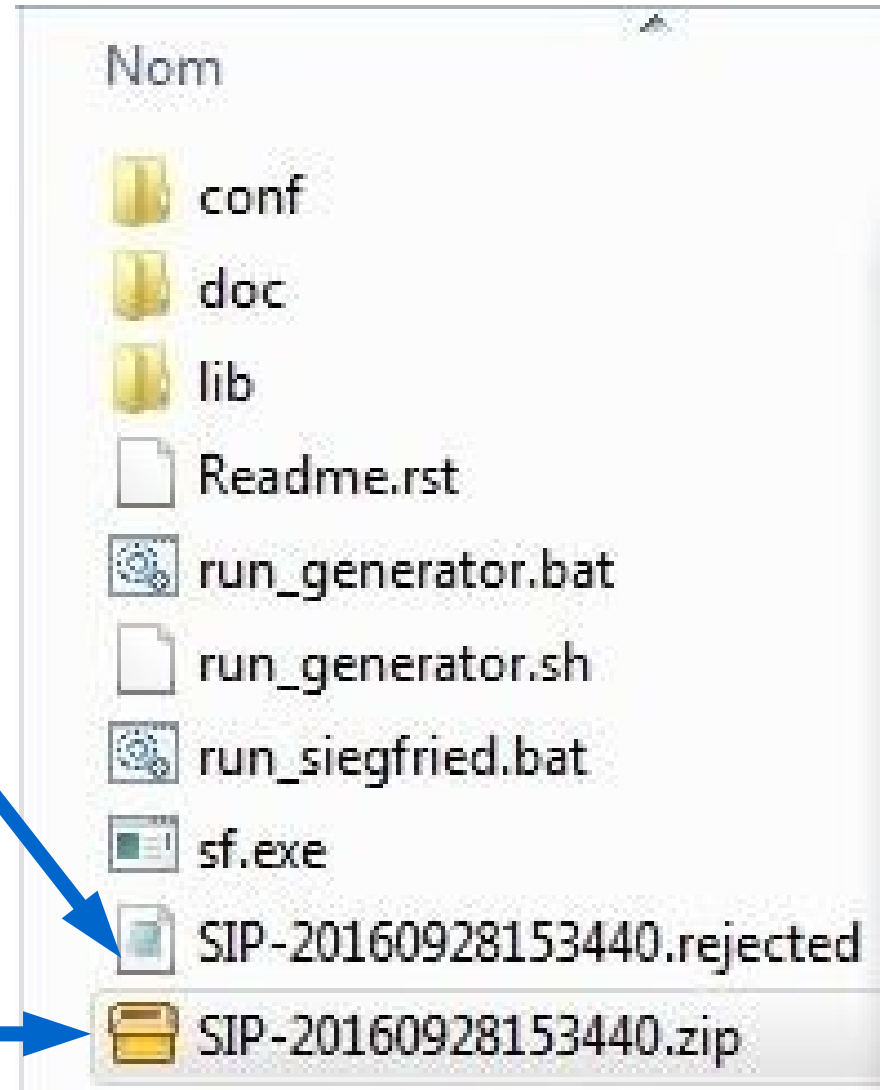
- Deux consoles s'ouvrent. La 2^e console trace les erreurs survenues pendant l'opération. Elle se ferme à la fin de l'opération et tapant sur la touche « entrée »



Étape 3 - où retrouver le SIP généré ?

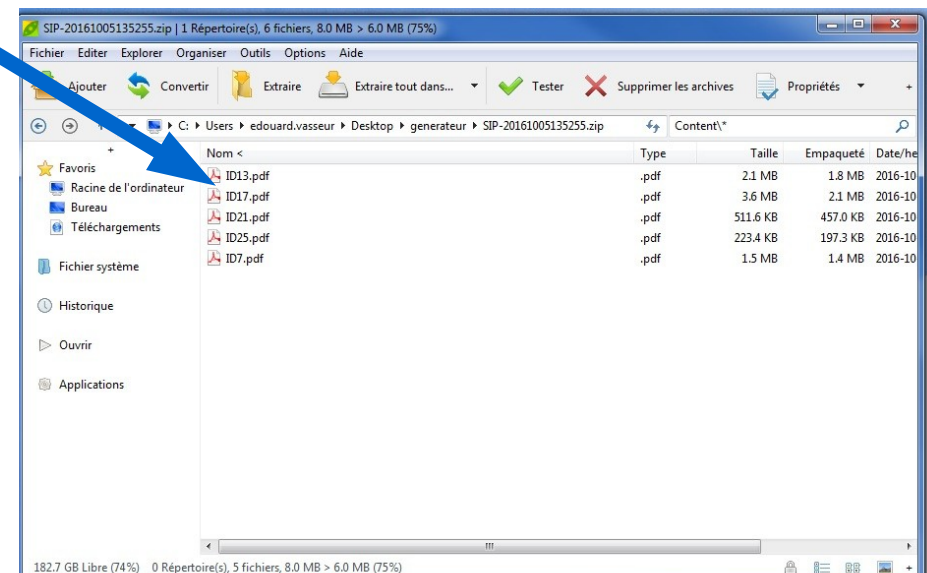
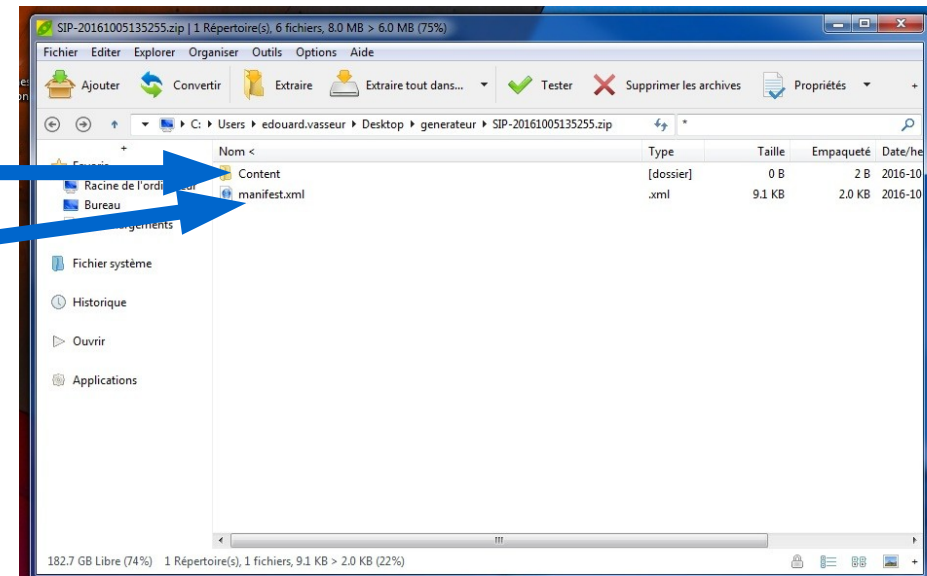
2 fichiers zip SIP ont été créés dans le répertoire contenant le générateur:

- Le 1^{er} « rejected » relève toutes les erreurs survenues lors de la création du SIP
 - *S'il n'y a pas d'erreur la taille du fichier est 0 Ko*
- Le 2^e est le SIP à proprement parler avec une extension .zip



Étape 4 - Consulter le SIP

- Ouvrir le zip et constater qu'il contient bien :
 - Un répertoire content
 - Un fichier manifest.xml qui correspond au bordereau
- Ouvrir le répertoire content :
 - Il comprend l'ensemble des fichiers de l'arborescence, renommés et à plat
- NB :
 - l'ancien nom des fichiers figure dans le bordereau comme métadonnée
 - le bordereau peut être extrait pour enrichissement (ex. avec un éditeur xml) et réinjecté dans le SIP ensuite (en supprimant la version d'origine)



Le SIP est prêt !

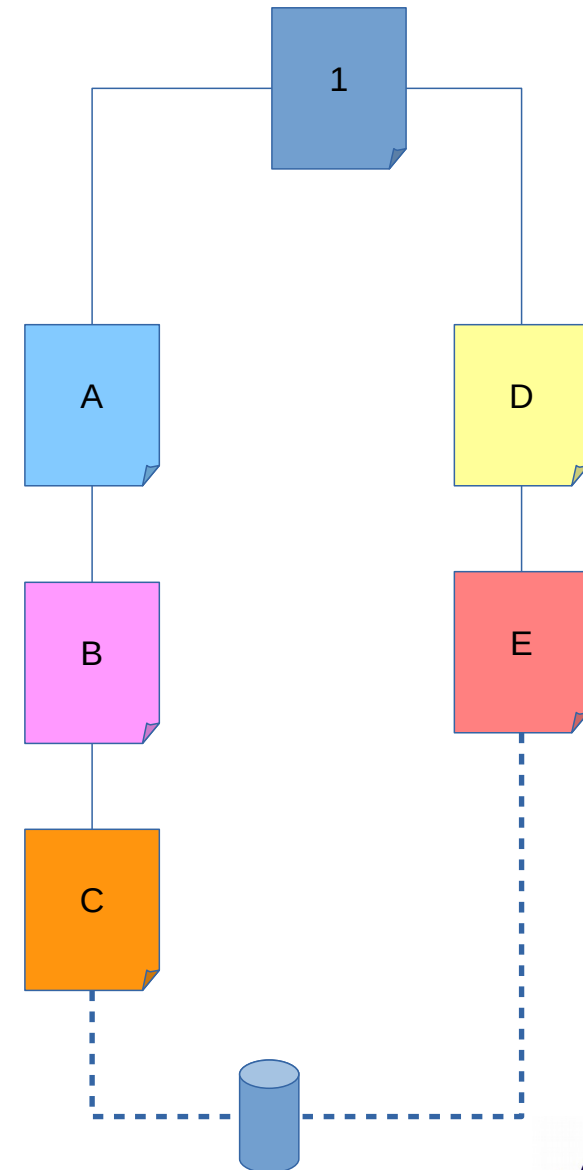
Fonctionnement avancé du générateur

Constituer des jeux de tests « avancés »

Organiser l'arborescence en fonction du comportement attendu

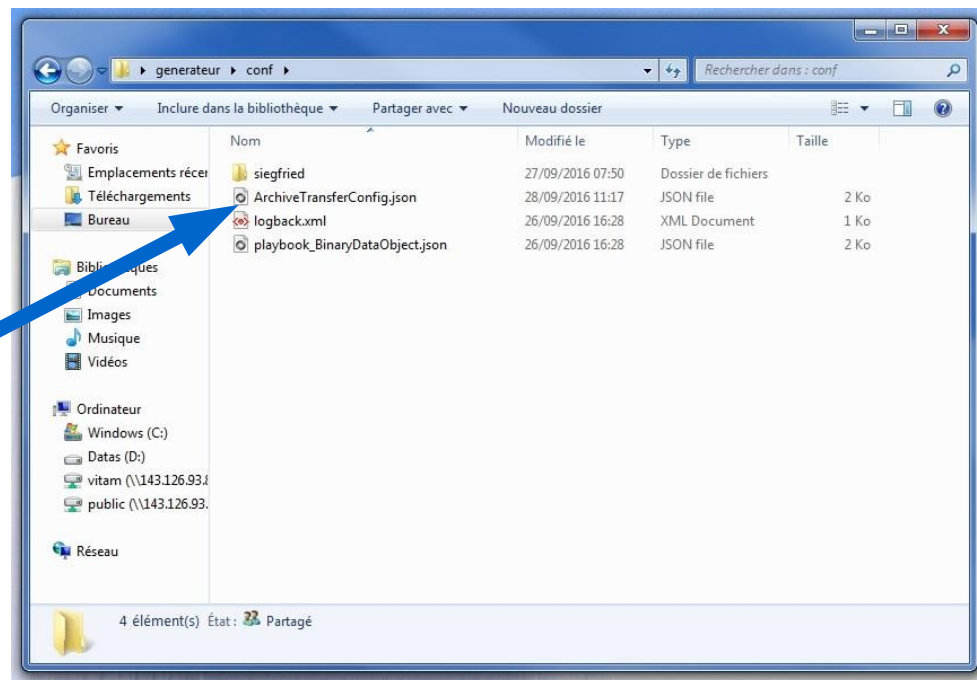
Ex. : avoir un fichier à plusieurs endroits dans l'arborescence

- Créer 1 répertoire comprenant 2 sous-répertoires (A et D)
- Prendre le 1^{er} sous-répertoire et y mettre 1 seul sous-sous-répertoire (B), dans lequel on crée un sous-sous-sous-répertoire (C) dans lequel on positionne le ou les fichiers
- Prendre le 2^e sous-répertoire et y mettre 1 seul sous-sous-répertoire (E), dans lequel on positionne le ou les mêmes fichiers que dans le sous-sous-sous-répertoire (C)



Paramétrage de l'en-tête du bordereau (1)

Il est possible de paramétrer l'en-tête du bordereau en utilisant le fichier ArchiveTransferConfig.json enregistré dans le répertoire conf



Paramétrage de l'en-tête du bordereau (2)

- Copier ce fichier dans le répertoire racine de l'arborescence
- Ouvrir le fichier .json avec un éditeur xml ou un éditeur de texte comme Notepad++
- Modifier les valeurs des champs qui intéressent l'utilisateur en respectant la ponctuation (guillemets) :
 - Comment : chaîne de caractère décrivant le transfert
 - MessagIdentifier : identifiant du transfert
 - ArchivalAgreement : contrat d'entrée utilisé
 - CodeListVersions : listes de codes utilisés dans le bordereau
 - ArchivalAgency : identifiant du service d'archives destiné à recevoir le transfert
 - TransferringAgency : identifiant de l'opérateur de transfert à l'origine du transfert
 - OriginatingAgencyIdentifier : identifiant du service producteur
 - SubmissionAgencyIdentifier : identifiant du service versant
- Indiquer les catégories de fichiers que l'utilisateur souhaite exclure du SIP (champ ignore_patterns)
- Sauvegarder les modifications

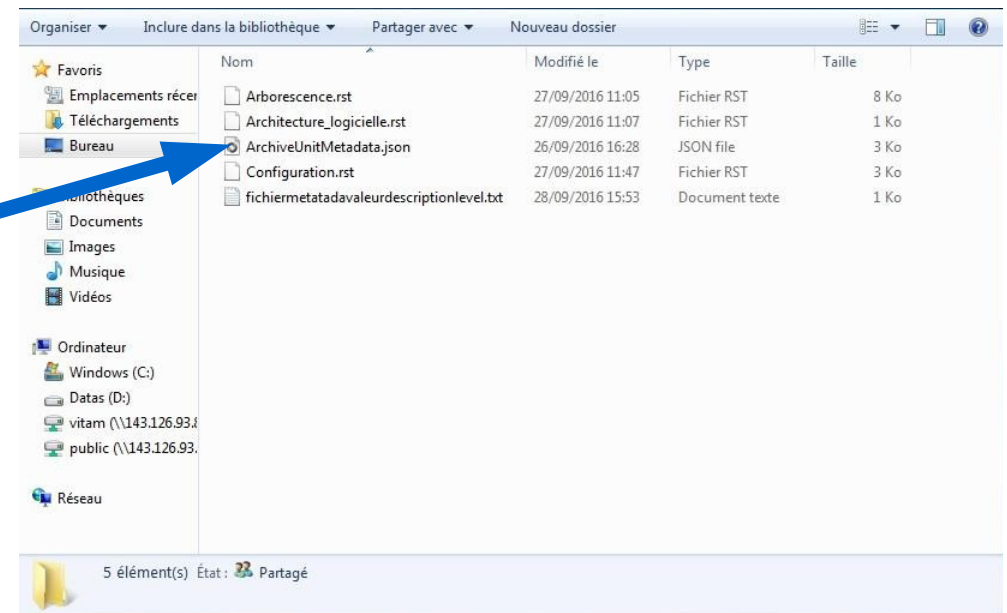
Paramétrage de l'en-tête du bordereau (3)

```
{
  "Comment" : "2eme SIP",
  "MessageIdentifier" : "MessageIdentifier0",
  "ArchivalAgreement" : "ArchivalAgreement0",
  "CodeListVersions" : {
    "ReplyCodeListVersion" : "ReplyCodeListVersion0",
    "MessageDigestAlgorithmCodeListVersion" : "MessageDigestAlgorithmCodeListVersion0",
    "MimeTypeCodeListVersion" : "MimeTypeCodeListVersion0",
    "EncodingCodeListVersion" : "EncodingCodeListVersion0",
    "FileFormatCodeListVersion" : "FileFormatCodeListVersion0",
    "CompressionAlgorithmCodeListVersion" : "CompressionAlgorithmCodeListVersion0",
    "DataObjectVersionCodeListVersion" : "DataObjectVersionCodeListVersion0",
    "StorageRuleCodeListVersion" : "StorageRuleCodeListVersion0",
    "AppraisalRuleCodeListVersion" : "AppraisalRuleCodeListVersion0",
    "AccessRuleCodeListVersion" : "AccessRuleCodeListVersion0",
    "DisseminationRuleCodeListVersion" : "DisseminationRuleCodeListVersion0",
    "ReuseRuleCodeListVersion" : "ReuseRuleCodeListVersion0",
    "ClassificationRuleCodeListVersion" : "ClassificationRuleCodeListVersion0",
    "AuthorizationReasonCodeListVersion" : "AuthorizationReasonCodeListVersion0",
    "RelationshipCodeListVersion" : "RelationshipCodeListVersion0"
  },
  "ArchivalAgency" : {
    "Identifier" : "Identifier4"
  },
  "TransferringAgency" : {
    "Identifier" : "Identifier5"
  },
  "ManagementMetadata.OriginatingAgencyIdentifier" : "Service_producteur",
  "ManagementMetadata.SubmissionAgencyIdentifier" : "Service_versant",
  "ignore_patterns" : ["Thumbs.db","pagefile.sys"]
}
```

Paramétrage des métadonnées d'un répertoire (1)

Il est possible de paramétrer l'indexation du répertoire en utilisant le fichier

ArchiveUnitMetadata.json enregistré dans le répertoire doc



Paramétrage des métadonnées d'un répertoire (2)

- Copier ce fichier dans le répertoire que vous voulez indexer
- Ouvrir le fichier .json avec un éditeur xml ou un éditeur de texte comme Notepad++
- Identifier les champs que vous voulez enrichir parmi la liste de tous les champs disponibles dans l'ontologie fournie par le SEDA
- Modifier les valeurs des champs qui vous intéressent en respectant la ponctuation (guillemets)
 - Voir slides correspondantes pour la liste des champs
 - Les champs DescriptionLevel et Title sont obligatoires
- Supprimer les champs que vous ne souhaitez pas voir apparaître dans l'indexation du répertoire
 - Voir slide correspondante pour les points d'attention
- Sauvegarder les modifications

Paramétrage des métadonnées d'un répertoire (3)

```
veUnitMetadata.json | ArchiveTransferConfig.json | ArchiveUnitMetadata.json |
{
  "Content" : {
    "DescriptionLevel" : "RECORD_GRP",
    "Title" : [ {
      "Value" : "Titre francais",
      "Lang" : "fr"
    }, {
      "Value" : "English title",
      "Lang" : "en"
    } ],
    "FilePlanPosition" : "Valeur de filePlanPosition",
    "SystemId" : "Valeur de SystemID)",
    "OriginatingSystemId" : "Valeur de OriginatingSystemId",
    "ArchivalAgencyArchiveUnitIdentifier" : "Valeur de archivalAgencyArchiveUnitIdentifier",
    "OriginatingAgencyArchiveUnitIdentifier" : "Valeur de originatingAgencyArchiveUnitIdentifier",
    "TransferringAgencyArchiveUnitIdentifier" : "Valeur de transferringAgencyArchiveUnitIdentifier",
    "Description" : [ {
      "Value" : "Description francaise",
      "Lang" : "fr"
    }, {
      "Value" : "English Description",
      "Lang" : "en"
    } ],
    "Type" : {
      "Value" : "Valeur du type",
      "Lang" : "fr"
    },
    "DocumentType" : {
      "Value" : "fr"
    },
    "Language" : "FR",
    "DescriptionLanguage" : "FR",
    "Status" : "Valeur de Status",
    "Version" : "Valeur de version",
    "Tag" : [ "XML Tag 1 (de type xml:token)", "XML Tag 2 (de type xml:token)" ],
    "Coverage" : {
      "Spatial" : [ {
        "Value" : "Valeur de Spatial",
        "Lang" : "fr"
      } ]
    }
  }
}
```

Paramétrage des métadonnées d'un répertoire (4)

Champs modifiables

- DescriptionLevel : niveau de description
 - *Voir slide correspondante pour la liste des valeurs*
- Title : titre
- FilePlanPosition : position dans le plan de classement
- SystemId : identifiant fourni par le SAE
- OriginatingSystemId : identifiant fourni par le système de production
- ArchivalAgencyArchiveUnitIdentifier : identifiant fourni par le service d'archives (ex. cote)
- OriginatingAgencyArchiveUnitIdentifier : identifiant fourni par le service producteur
- TransferringAgencyArchiveUnitIdentifier : identifiant fourni par le service versant
- Description : présentation détaillée du contenu
- Type : type d'information au sens de l'OAIS

Paramétrage des métadonnées d'un répertoire (5)

Champs modifiables

- DocumentType : type de document
- Language : langue des archives
- DescriptionLanguage : langue des descriptions
- Status : statut du document
- Version : version du document
- Tag : indexation
- Coverage : couverture (Spatial : géographique ; Temporal : chronologique ; jurisdictional : administrative)
- OriginatingAgency : identifiant du service producteur
- SubmissionAgency : identifiant du service versant
- Writer : rédacteur du document
- Source : pour les documents numérisés, référence au document source
- Event : description d'un événement survenu pendant le cycle de vie du document
- Gps : coordonnées géographiques définies par l'utilisateur

Paramétrage des métadonnées d'un répertoire (6)

- Valeurs possibles pour le champ Niveau de description
 - Fonds : mettre FONDS
 - Sous-fonds : mettre SUBFONDS
 - Classe : mettre CLASS
 - Collection : mettre COLLECTION
 - Série : mettre SERIES
 - Sous-série : mettre SUBSERIES
 - Groupe d'archives : mettre RECORD_GRP
 - Sous-groupe d'archives : mettre SUB_GRP
 - Dossier : mettre FILE
 - Pièce : mettre ITEM

Paramétrage des métadonnées d'un répertoire (7)

Points d'attention sur la suppression des champs

- Toujours mettre une virgule à la fin de chaque champ, sauf à la fin du dernier champ
- les champs Title et Description peuvent avoir plusieurs valeurs. Il faut vérifier qu'il y a bien le bon nombre de caractères spéciaux utilisés pour délimiter le champ au moment de la suppression

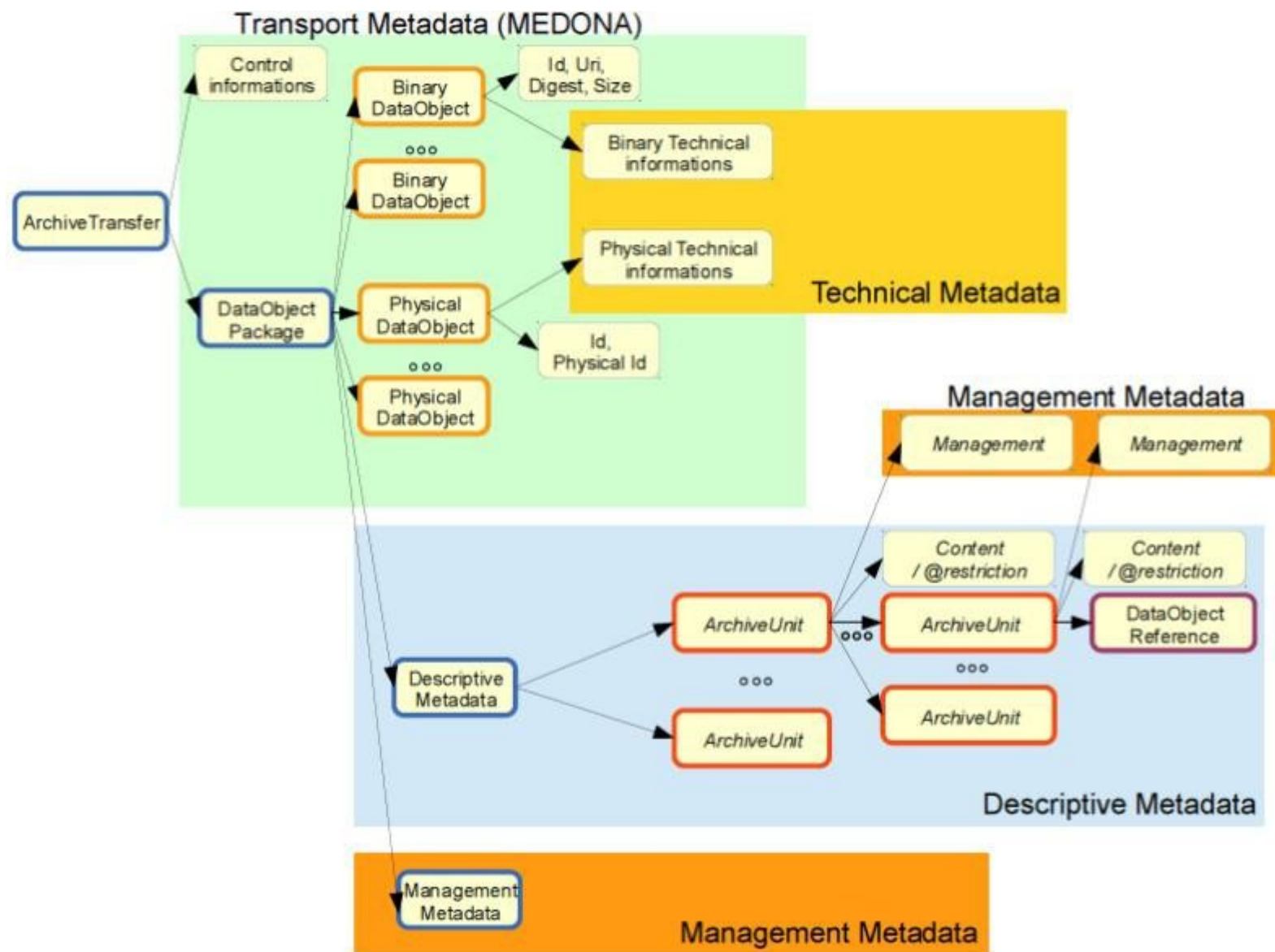
Paramétrage des métadonnées d'un répertoire (8)

Points d'attention sur la suppression des champs - exemple

```
{ "Content" : {  
  "DescriptionLevel" : "FILE",  
  "Title" : [ {  
    "Value" : "Documentation sur le standard METS  
  } ],  
  "FilePlanPosition" : "3.1.1.",  
  "OriginatingSystemId" : "24561",  
  "Description" : [ {  
    "Value" : "Documentation récupérée sur le site de la Bibliothèque du Congrès"  
  } ],  
  "Language" : "EN",  
  "OriginatingAgency" : {  
    "Identifier" : {  
      "Value" : "CodeVitam"  
    }  
  },  
  "StartDate" : "2016-09-26T15:34:08.284+02:00",  
  "EndDate" : "2016-09-26T15:34:08.284+02:00"  
}
```

Annexe sur le SEDA

Structuration du bordereau (message de type ArchiveTransfer)



Vocabulaire

- <DataObjectPackage> : englobe tous les objets et leurs MD de description et de gestion
- <BinaryDataObject> : objet correspondant à des fichiers binaires
- <PhysicalDataObject> : objet correspondant à quelque chose de physique (un carton, un cd-rom, etc.)
- <DataObjectGroup> : groupe d'objets données
- <DataObjectGroupId> : identifiant d'un groupe d'objets données
- <DataObjectVersion> : usage/version de l'ensemble intellectuel
- <Uri> : chemin permettant d'accéder au fichier
- <MessageDigest> : empreinte du fichier
- <Descriptive Metadata> : MD de description
- <Management Metadata> : MD de gestion
- <ArchiveUnit> : unité d'archives
- <Content> : contenu de l'unité d'archives
- <DataObjectReference> : référence interne à un objet-donnée ou à un groupe d'objet-données