

Table des matières

Index

4

Contexte

10

Objectifs

11

Contenu de la prestation

11

Caractéristiques du système final

12

Convivialité

12

Souplesse

12

Réutilisation

12

Utilisation des normes

12

Intégration

12

Saisie déconnectée

12

Description de la plate-forme de production

13

Description du mode de paramétrage de la plate-forme de production

13

Paramétrage de la saisie

13

Les liens	15
Gestion de l'intégrité des liens	15
Principe de la saisie	16
Aides à la saisie	16
Affichage des objets externes	16
Mode de rappel	16
Traitement par lot	17
Validation	17
Gestion des utilisateurs / saisie simultanée	18
Les tableaux	18
Gestion des évolutions de schéma.....	19
Gestion des objets multimédia	19
Géo-référencement.....	19
Gestion des versions des modèles	19
Méta-données système	20
Importation/Exportation	21
Importation	21
Exportation	21
Situation du noyau par rapport aux futurs développements	
22	
Les logiciels à fournir dans le cadre de la prestation	22
Systèmes cibles	22
Serveur	22
Clients	23
Performances	23
Reprise de l'existant	24
Délivrables	24
Formation	25
ANNEXE AU CAHIER DES	
CHARGES	26
1.1	Tests de validation à fournir par le prestataire
27	
Dossier d'archives (DAF)	27
Dossier d'artiste (DAP)	27
Dossiers d'artistes	27
Dossier de l'Inventaire Général	28

Dossier d'œuvre	30
1.2Méthode d'identification des dossiers et des fragments	31
1.3 Description des contrôles	32
1.4Modélisation des liens	33
1.5 Importation de documents	36
1.5.1 Règles standards pour l'importation normale de documents	36
1.5.2 Règles spéciales d'importation pour la reprise de l'existant	43

Index

A

accès 12, 18
adressage 15
aides à la saisie 13
annexe 24
API 16
ascii délimité 20
attribut 16, 21
attributs 14

B

base de données 13
bases 13
boutons 16

C

candidats 17
cartes 19
code 20
cohérence 12, 14
complet 18
constantes 16
contrôle 13, 14
contrôles 12, 13, 17
conversion 24
coordonnées 19
corpus 21
créateur 18

D

déplacement 15
diffusion 18
DOSI 8, 20
DTD 8

E

écrans de saisies 13
éditeur de texte 16
Élément 8
élément 16
entités 21
environnement graphique 12
erreur 15
Exportation 21
exportation 18
exporter 12

F

feuille de style 8, 13, 21
fonds 19
formel 17
formulaire 14
formulaires 13
fragment 8, 14, 15, 16

G

génération 13
géo-référencement 12, 13, 19

H

humain 17

I

icône 16
id 20
identifiant 20, 22
illustration 13, 19
image 19
Importation 20, 21
importation 13
indexation 13
Instance 8
intégrité 12, 18
interactif 15

J

JAVA 23

L

lien 13
liens 15, 18
liens groupés 15
liens simples 15
listes déroulantes 16
logiciels 22
longueur 14
lot 17

M

masque de saisie 22
menus 16
Méta-données 8, 20
méta-données 17
mosaïque 19
multimédia 15

N

notice 8, 12
numériques 17
numéro de fonds 20

O

objet 16, 19
objets 15, 21
objets multimédia 12
onglets 16
opérateurs 17
OVIDE 17

P

papier 22
Paramétrage 13
paramétrage 13

paramètres 21
plein texte 16
portable 12
prestataire 17, 22
producteur 20
propriétaire 18
publication 22

R

racine 20
recherche 13, 15, 16, 18
récolement 21
recommandations 12
référence 22
référencement 15
règles 21, 24
relations transversales 13
reprise 22, 24
responsable 18

S

saisie 12, 13
schéma de données 8
Schéma XML 8
sémantique 17
session 18
SGBD 22
simultanée 18
statut 17
stockage 13
SVG 8, 19
syntaxe 21

T

table 15
tableaux 18
thesaurus 16
transformation 21
transformations 19
typographie 17
typologie 14

U

utilisateurs 13, 18

V

Validation 9, 17
validation 12, 18
vecteur 19
vérification 18
verrouillage 12
vidéo 19
visualiseur 19
vocabulaire contrôlé 16
vue arborescente 16

X

XSL 9, 21, 22
XSLT 9

1.1 GLOSSAIRE

Base de documents	collection de documents produits avec le même schéma de données
Document	instance d'un schéma de données
Dossier électronique	ensemble documentaire constitué d'un document principal et de documents annexes, et pouvant inclure des objets multimédia
DTD	Data Type Définition : une DTD contient les spécifications formelles de structuration de documents au format XML.
DOSI	Département de l'Organisation et des Systèmes d'Information placé sous la direction de l'administration générale du MCC
Élément	Un élément est une unité de structuration de document déclarée dans un schéma de données.
feuille de style	Une feuille de style détermine l'apparence et la mise en forme d'éléments dans un ensemble de documents qui la référencent. Les documents XML ne contiennent en principe aucune information de présentation des données à l'affichage. Un document XML peut spécifier une feuille de style pour l'affichage de ses données, dans un navigateur par exemple.
fragment	Partie d'un document constituant une entité logique réutilisable. Un fragment est constitué d'un élément racine et de sous-éléments.
Internet	Réseau informatique s'appuyant sur le standard de fait TCP/IP mis au point par les USA dans les années 1960
Instance	Production d'un document selon un schéma de données.
MCC	Ministère de la Culture et de la Communication
MRT	Mission de la recherche et de la technologie (placée sous la Direction de l'administration générale du Ministère de la Culture et de la Communication)
Méta-données	Données décrivant des données. Un nom d'élément ou de champ, par exemple.
notice	unité documentaire issue d'une base de données, et structurée en une liste de champs.
schéma de données	spécification formelle de la structure d'un document sous forme d'arbre.

dimensions exprimé en XML.

Validation

D'un point de vue formel, vérifier qu'un document est conforme à la structure et à la syntaxe définie dans un schéma de données.

W3C

World Wide Web Consortium : consortium international ayant pour mission d'établir des recommandations de formalisation des structures d'échange de données sur le WEB

WEB

toile d'interconnexion mondiale de machines informatiques s'appuyant sur le réseau Internet

XML

Extensible Markup Language : Langage de marquage de documents.

Contexte

Depuis plus de cinq ans, le ministère de la culture a réalisé plusieurs programmes de recherche visant à définir son futur système d'information documentaire. Ces recherches ont été menées aussi bien au plan national qu'au plan international, notamment dans le cadre du projet européen Aquarelle.

La principale conclusion de ces projets de recherche est que toute la documentation scientifique produite par les services du ministère doit être électronique et stockée au format XML.

De plus, il a été entrepris des études complémentaires afin de définir quelles étaient les schémas XML les plus appropriés pour chaque type de documents. Aujourd'hui, il existe un schéma XML pour chaque domaine d'activité du ministère (bibliothèques, archives, musées, monuments, etc.).

Pendant ces expérimentations, des utilisateurs du ministère ont pu utiliser différents outils pour produire des documents au format XML : éditeurs de documents SGML adaptés à XML (ArborTex, Griff), traitement de textes (WordPerfect), logiciel spécifique (système développé à partir du SGBD Access).

Le présent cahier des charges s'appuie sur l'ensemble de ces travaux. Il tient compte également des développements informatiques réalisés par le DOSI (département de l'organisation et des systèmes d'information du ministère), ainsi que ceux mis en place par certaines DRAC (Direction Régionale des Affaires culturelles).

Les bases de données documentaires nationales qui sont actuellement accessibles en ligne sur Internet sont exploitées à l'aide du logiciel MISTRAL de la société BULL. Elles devront à terme être publiées sous forme de dossiers électroniques au format XML. Lorsque le système de production de dossiers XML, objet du présent cahier des charges, sera opérationnel, une reprise de l'antériorité sera entreprise pour la migration des données des anciennes bases vers le nouveau système. Parallèlement à cette phase de reprise, les dossiers créés directement sur la plate-forme de production pourront être extraits sous forme de notices documentaires classiques afin de mettre à jour les bases de données existantes. Les bases dont la migration sera achevée pourront faire l'objet d'un nouveau mode de publication reposant exclusivement sur les dossiers électroniques. Le futur système de publication en ligne de dossiers électroniques, non défini à ce jour et non inclus dans le présent cahier des charge, doit être étudié prochainement. Les services qui souhaiteraient publier rapidement en ligne des dossiers électroniques produits à l'aide de la plate-forme de production pourront utiliser des outils de publication Internet existants. (voir avec la Mission de la Recherche et de la Technologie)

Objectifs

L'objectif du MCC est de mettre en place un système de production, de gestion et de diffusion de documents électroniques, basé sur la norme internationale XML.

Le système comprend les éléments suivants :

- une plate-forme de production paramétrable, permettant la saisie, le stockage et l'indexation des documents,
- un système d'archivage au format XML indépendant,
- un module de reprise de l'existant (conversion de notices classiques ou reprise de documents XML existants),
- un module de diffusion multi-supports (Intranet/Internet, fichiers, papier).

Contenu de la prestation

La présente consultation a pour objet le développement :

- de la plate-forme de production,
- la reprise et l'adaptation du système de contrôle de documents et de gestion de thésaurus développé par le MCC, pour son intégration avec la plate-forme,
- du système d'archivage,
- du module de reprise de l'existant.

Le prestataire doit fournir les outils permettant le paramétrage du système, c'est-à-dire :

- un éditeur de schéma XML ;
- un éditeur de feuilles de style (XSL/XSLT) ;

Le module de diffusion multi-supports sera réalisé ultérieurement.

Caractéristiques du système final

Convivialité

L'application devra être intégrée dans un environnement graphique et prendra en compte certaines particularités dues à la nature des informations manipulées ; il conviendra par exemple d'éviter la saisie répétitive d'informations.

Souplesse

L'application devra permettre aux chercheurs de rédiger leurs dossiers sans contraintes. En effet, bien qu'un schéma XML corresponde à la structure finale d'un dossier, les chercheurs doivent pouvoir commencer la saisie des données par des blocs de données indépendants (références bibliographiques, illustrations, etc.) et ne doivent pas subir l'ordre imposé par les schémas XML.

Réutilisation

Pour pouvoir être réutilisé par tous les services du MCC, l'outil de production devra rester le plus générique possible et ne pas se focaliser sur les aspects spécifiques d'un schéma XML d'un service. Les différents services du ministère vont, dans un futur proche, disposer de leur propre schéma XML; on aura ainsi, à terme, une dizaine de schémas de dossiers XML au MCC et l'outil de production devra être capable de s'adapter à ces structures de données.

L'ajout d'un nouveau schéma XML ne devra donc pas engendrer une refonte du système ; seul un paramétrage des différents composants sera à réaliser.

Les différents contrôles de validation doivent également être paramétrables pour s'adapter aux contraintes propres à chaque type de document.

Utilisation des normes

Les développements objets du présent cahier des charges devront tenir compte des recommandations du W3C (Xschéma, Xpath, Xlink, XSL/XSLT, SVG, etc.).

Intégration

Le futur outil de production doit permettre à l'utilisateur de déclencher des contrôles d'intégrité des données (contrôles lexicaux, syntaxiques et de cohérence), d'aller chercher des illustrations dans un fonds d'objets multimédia, d'insérer des notices sur des ouvrages à partir d'une base bibliographique ou d'accéder à un système de géo-référencement pour récupérer les coordonnées d'un édifice automatiquement, tout ceci depuis une seule et même interface.

Saisie déconnectée

Le système doit offrir la possibilité de saisir des documents sur des stations portables et de les réintégrer dans la plate-forme de production.

Un système de verrouillage doit permettre aux utilisateurs, après création d'un document sur la plate-forme de production, d'exporter ce document et d'en verrouiller l'accès jusqu'à son rapatriement.

Description de la plate-forme de production

La plate-forme de production doit pouvoir gérer plusieurs types de documents, chaque type de document étant lié à un schéma XML. Un mécanisme de lien permettra de mettre en relation les documents entre eux. Certains documents constitueront l'ossature de dossiers complexes. D'autres documents dits individuels viendront se rattacher aux dossiers par l'intermédiaire de liens mis en œuvre par les utilisateurs. Les documents individuels pourront ainsi constituer des bases partageables pour l'ensemble des autres documents de la plate-forme (ex : base auteur, base bibliographique, base d'illustration, etc.).

En dehors des relations liées à la structure des dossiers, d'autres relations transversales pourront être mise en œuvre (relation entre parties de documents permettant d'assurer des renvois entre dossiers ou entre parties d'un même dossier).

La plate-forme de production comprend les fonctions suivantes :

- saisie par une interface graphique des éléments constitutifs des documents,
- importation de données pour intégration dans des documents,
- mise en relation de documents,
- géo-référencement des documents,
- illustration des documents,
- aides à la saisie (accès aux thésauri, gestion de constantes, etc),
- contrôle des documents (existant à adapter),
- gestion des utilisateurs,
- stockage des documents dans une base de données ;
- indexation et recherche de documents dans la base de données ;
- génération de dossiers ou de documents au format XML avec ou sans application de feuille de style XSL/XSLT (exportation);

Lors de la saisie ou de la visualisation des documents, les utilisateurs doivent pouvoir visualiser les données dans des formulaires, dans des documents mis en page, ou sous forme arborescente.

Description du mode de paramétrage de la plate-forme de production

Paramétrage de la saisie

Les informations à saisir font l'objet d'une description formalisée dans un schéma de données . La plate-forme repose sur l'utilisation de formulaires de saisie pour certaines parties des schémas.

La plate-forme est paramétrable en fonction de différents schémas de données et de leurs évolutions. Ainsi, à partir de chaque schéma de données, la plate-forme devra être configurée au niveau :

- du schéma de stockage interne des données (si nécessaire¹) ;
- des écrans de saisies et des aides associées,
- des contrôles

Les éléments nécessaires au paramétrage de la saisie sont décrits dans des fichiers externes associés à chaque schéma de données.

Le principe de fonctionnement du paramétrage repose sur la définition de formulaires de saisie pour certaines parties des schémas.

Le fichier de paramètre définit pour chaque formulaire la suite d'éléments entrant dans sa composition. Lorsqu'un élément possède des attributs, ces attributs peuvent accompagner l'élément dans le formulaire. Les éléments ou attributs utilisés dans un formulaire doivent faire partie d'un seul fragment.

La définition du formulaire contient les informations suivantes :

- nom de l'élément dans le schéma;
- nom externe de l'élément (libellé à faire apparaître dans le formulaire) ;
- informations sur l'aide à la saisie (liste de valeurs, note explicative)

Les paramètres suivants, intégrés au schéma de données, sont pris en compte pour la génération des formulaires et la saisie:

- typologie du champ (numérique, numérique + décimales, date, alphabétique pur, alphanumérique) ;
- longueur minimale de la donnée à saisir ;
- longueur maximale de la donnée à saisir ;

Les paramètres concernant le contrôle sémantique des éléments de formulaire sont intégrés au système de validation des documents (voir chapitre *Validation – validation sémantique*).

Un mécanisme permet de contrôler la cohérence des formulaires.

Les liens

La modélisation des liens s'appuie sur la norme Xlink. Le mécanisme de gestion des liens (saisie, stockage, restitution) devra permettre d'utiliser les différents types de liens utilisés actuellement dans les schémas de données du Ministère (DTD EAD des Archives, DTD classeur de l'Inventaire et DTD des dossiers d'artiste). Un paramétrage permettra d'identifier les éléments permettant de mettre en œuvre des liens, de gérer dynamiquement la liste des éléments ou attributs optionnels, de définir des libellés clairs pour la présentation des éléments ou attributs lors de la saisie.

Le système doit gérer des liens de natures différentes

- Les liens simples entre deux entités documentaires ou multimédia
- Les liens groupés entre plusieurs entités documentaires et multimédia (voir schéma EAD)

En plus des informations liées à la norme Xlink, Les informations suivantes doivent prises en compte dans la gestion des liens :

- la date de création ou de dernière modification du lien;
- l'identification du créateur du lien (identifiant d'une personne physique) ;
- un commentaire (optionnel)

Mise en œuvre

L'utilisateur dispose d'un outil interactif lui permettant de désigner les objets source et cible et de compléter les informations nécessaires à la description du lien. Un interface de recherche de documents, de fragment de document ou d'objets multimédia est intégré à l'outil.

Les informations concernant les liens ne figurent pas directement dans les documents. Ces informations sont stockées dans une table indépendante afin de permettre une mise à jour plus facile en cas de déplacement des objets externes associés aux dossiers (changement de répertoire, changement de serveur de fichier, etc). Le système de référencement par entités est utilisé pour l'adressage des objets impliqués dans les liens.

Gestion de l'intégrité des liens

Le système doit prendre en compte la gestion de l'intégrité des liens.

Le principe de base est qu'un élément lié à un autre ne peut être supprimé (on supprimera d'abord le lien).

Lors de la suppression d'un dossier il sera proposé de supprimer tous les éléments qui lui sont liés, à l'exclusion des éléments impliqués dans un lien avec un élément appartenant à un autre dossier.

Lors de la suppression d'un objet externe, il sera proposé de supprimer le lien vers cet objet ou de rechercher dans l'environnement local le nouvel emplacement de l'objet..

Lorsqu'un document est ouvert, les liens de ce document vers des objets externes sont vérifiés . En cas d'erreur (objet absent ou accès impossible) l'utilisateur est prévenu et doit décider de la suppression ou non du lien.

Principe de la saisie

Deux modèles de saisie sont proposés à l'utilisateur. Le premier modèle est basé sur une présentation arborescente des documents (modèle proposé habituellement dans les éditeurs XML). Le deuxième modèle permet de saisir certaines parties de document en utilisant des formulaires prédéfinis. La présentation des formulaires doit être ergonomique et peut reposer sur la définition d'onglets, de menus « surgissants », de listes déroulantes.

Le fait de sélectionner un formulaire :

- appelle le formulaire de saisie du fragment cible ;
- charge ce formulaire dans le cas où le fragment est déjà rempli.

Pour chaque formulaire, l'utilisateur dispose de boutons permettant :

- de supprimer le fragment (sous réserve du respect des contraintes d'intégrité) ;
- de modifier le fragment;
- de saisir un nouveau fragment;
- d'appeler un nouveau formulaire pour les éléments contenant d'autres éléments ;
- de revenir au formulaire précédent après validation de la saisie (pour les formulaires appelés à partir d'un autre formulaire)
- de relier le fragment à une entité interne ou externe (autre fragment ou objet multimédia) si le schéma de données l'autorise et selon les modalités décrites à l'article 8.

Au lancement d'une session, l'utilisateur peut choisir de débiter la saisie soit par le système d'onglets, soit par un modèle plus classique avec vue arborescente et éditeur de texte.

Aides à la saisie

Saisie répétitive

Les utilisateurs pourront définir des valeurs constantes à insérer systématiquement dans certains éléments lors de la création d'un nouveau dossier. Ces valeurs constantes seront paramétrables en ligne (changement en cours de saisie).

Utilisation de vocabulaire contrôlé

Lors de la saisie d'éléments à vocabulaire contrôlé, les utilisateurs pourront faire appel à des listes simples (stockées dans des fichiers externes ou spécifiées dans le schéma de données utilisé) ou au gestionnaire de thesaurus fourni par le Ministère sous forme d'API (Interface programmatique d'application), avec présentation arborescente des thésauri à prévoir.

Affichage des objets externes

Les objets externes reliés à un document apparaîtront sous forme d'icône. L'utilisateur pourra cliquer sur l'icône pour consulter l'objet (activation d'un éditeur propre au type d'objet relié). Dans le cas d'un document, une nouvelle fenêtre s'affichera avec le document cible.

Mode de rappel

L'utilisateur dispose d'un interface de recherche des documents XML utilisant les critères suivants:

- recherche sur tout élément et/ou attribut faisant partie du schéma,
- recherche plein texte (sans préciser d'élément ou d'attribut),

- utilisation des opérateurs et/ou/sauf/proximité,
- test sur valeurs numériques et dates (=, >, <),
- possibilité d'interroger en tenant compte ou non de la typographie.

Les méta-données système (voir chapitre *Métadonnées Système* plus bas) peuvent être utilisées pour la recherche (ex : date de création d'un dossier, nom du créateur) ainsi que le statut des documents (validé / non validé par un responsable).

Traitement par lot

L'utilisateur pourra sélectionner un corpus de documents et demander la modification d'un élément ou d'un attribut sur ce corpus. La modification pourra s'appliquer à :

- une chaîne de caractères, un mot, un groupe de mots ou le contenu de l'élément ou de l'attribut,
- la première, la dernière ou toutes les occurrences de la séquence à modifier.

Validation

Un système de validation doit permettre de déterminer si les documents produits sont terminés ou en cours d'achèvement afin de les déclarer candidats ou non pour des traitements avals (indexation, diffusion). Dans cette validation il faut prendre en compte trois aspects :

- l' aspect formel
- l' aspect sémantique
- l' aspect humain.

Chaque document possède un statut qui permet de savoir s'il est dans l'état saisi (niveau formel validé), contrôlé (niveau sémantique validé) ou valide (validation par un responsable).

- Validation formelle

Cette validation s'appuie sur les spécifications introduites dans les schémas XML (structure et type).

- Validation sémantique

Le modèle de contrôle actuellement en vigueur dans l'outil OVIDE (Outil de Vérification d'Intégrité des Documents Electroniques) utilisé par le Ministère pour la validation des notices documentaires versées dans les bases de données nationales sera repris et adapté à la plate-forme de saisie des dossiers électroniques (voir les détails dans le chapitre Description des *contrôles* en annexe). Ce modèle s'appuie sur différents types de contrôles :

- Les contrôles de syntaxe (niveau élément ou attribut)
- Les contrôles de vocabulaire (consultation de thesaurus)
- Les contrôles de cohérence (niveau document)
- Les contrôles de liens

Un document sera remis au prestataire pour l'aider dans la reprise et l'adaptation des contrôles actuels à la plate-forme de production des dossiers électroniques (liste des règles, fonctions ou macro-procédures utilisées avec leur description et leur syntaxe, description du mécanisme d'appel des contrôles).

Mise en œuvre

Les contrôles sont activés de trois façons différentes :

- pendant la saisie d'un élément (demande explicite)

- en sortie d'élément ou d'attribut (à posteriori, contrôle unitaire)
- en sortie de dossier (contrôle d'intégrité complet sur le dossier).

Parallèlement à la reprise des contrôles existants, un système de spécification et d'exécution de contrôle de cohérence à base de schémas XML sera réalisé. Les schémas XML de spécification des contrôles s'appuieront sur la syntaxe indiquée dans le *chapitre Description des contrôles d'intégrité* en annexes - rubrique *implémentation future*).

Les contrôles de cohérence incluent la vérification d'intégrité des liens établis entre des documents et des objets externes. Lorsqu'un objet multimédia associé à un document sera déplacé ou supprimé, l'utilisateur devra en être averti au moment du contrôle d'intégrité complet sur le dossier.

· Validation humaine

Les documents devront être visés par un responsable pour être considérés comme valide. L'utilisateur aura la possibilité de rechercher des lots de documents et de consulter une liste résultat incluant leur statut. Ce module de recherche pourra être le même que celui utilisé pour la recherche de document lors de la saisie. Des traitements tels que l'exportation et la diffusion devront prendre en compte le fait qu'un dossier a été validé ou non par un responsable (autorisation de diffusion).

Gestion des utilisateurs / saisie simultanée

Saisie simultanée

Un même dossier pourra être saisi par plusieurs personnes simultanément. Un mécanisme de verrouillage permettra à un utilisateur d'être averti en cas d'utilisation du même fragment de données par un autre utilisateur.

Gestion des utilisateurs

Les dossiers créés appartiennent à l'utilisateur qui les a créés. Celui-ci peut être déclaré le seul habilité à les modifier. Pour la lecture trois cas sont envisagés :

- soit le dossier n'est visible que par son propriétaire/créateur,
- soit il est visible par tous
- soit il est visible par un groupe.

Une gestion de session devra être mise en place pour assurer un accès aux documents à travers un profil utilisateur (gestion des accès en lecture/écriture, accès réservé pour certains utilisateurs à certains documents) et gérer leur validation (notion de rôle associée à chaque utilisateur).

Les tableaux

Les services du Ministère utiliseront le même schéma pour modéliser les tableaux incorporés aux dossiers électroniques. Le prestataire devra développer un formulaire de saisie de tableaux conforme à ce schéma XML pour permettre aux utilisateurs d'incorporer ce type d'objet à leurs dossiers. Il sera possible, à travers l'éditeur de dossier de modifier les dimensions du tableaux au cours du temps (ajout, suppression de colonnes et de lignes).

Le modèle choisi pour la gestion des tableaux devra respecter les modèles utilisés dans les schémas de

données fournis par le Ministère (DTD EAD, DTD des classeurs de l'Inventaire Général, DTD des dossiers d'artiste).

Gestion des évolutions de schéma

En cas d'évolution d'un schéma de données, la plate-forme de production intègre une procédure capable d'opérer une sélection sur une collection de documents existants (sans téléchargement de la base) et d'appliquer les transformations indiquées dans une feuille XSLT. Cette procédure est activée sur demande.

Gestion des objets multimédia

Pour faciliter l'illustration des dossiers, le système devra comporter un visualiseur simplifié. Ce visualiseur devra permettre de :

- rechercher un ou plusieurs objet multimédias dans l'environnement du PC qui l'utilise (notamment le lecteur de CD),
- visualiser les objet multimédia sous forme de mosaïque (icônes banalisés pour les séquences son et vidéo) et en consultation (affichage grand écran pour une image, exécution de la séquence pour un son ou un film vidéo),
- sélectionner un ou plusieurs objets multimédias,
- créer le lien entre les objets multimédias sélectionnés et le dossier ouvert.

Géo-référencement

Un outil de géo-référencement intégré à la plate-forme de production permet d'alimenter les éléments de localisation des dossiers électroniques à partir d'objets sélectionnés sur des cartes (coordonnées ponctuelles ou emprise au sol d'édifices, tracé de réseau viaire, localisation d'objets, etc.).

Cet outil permet de charger des fonds cartographiques géoréférencées (image ou vecteur) et des localisations d'objets prédéfinis. Les données cartographiques fournies par le Ministère seront soit dans un format d'exportation du logiciel ArcView soit au format SVG (sous réserves).

Gestion des versions des modèles

La plate-forme assurera la gestion des versions de documents en fonction des versions de schémas (écri-

ture dans les documents des versions de schéma, lecture d'un document en fonction du schéma indiqué par la numéro de version).

Méta-données système

Certaines informations sont nécessairement associées à chaque document produit - en dehors des informations incluses dans les schéma de données - pour permettre au système d'assurer les fonctions décrites dans ce cahier des charges (par exemple la gestion des utilisateurs). Ces informations que l'on peut qualifier de *méta-données système* et qui doivent être prises en compte tant au niveau de la création que de la restitution des documents XML (à inclure dans les critères de recherche possible et à faire apparaître dans les documents exportés) sont indiquées dans la liste suivante (non exhaustive) :

1. IDENTIFIANT DU DOCUMENT
2. DATE DE CREATION DU DOCUMENT (AAAA/MM/JJ)
3. DATE DE DERNIERE MODIFICATION DU DOCUMENT (AAAA/MM/JJ)
4. ORGANISME PRODUCTEUR
5. CRÉATEUR (voir Gestion des utilisateurs)
6. RESPONSABLE (voir Principe de la saisie/Validation)
7. CONTROLE D'ACCES ET COMMUNICABILITE (voir Gestion des utilisateurs/Principe de la saisie/Validation)
8. LANGUE-DU-DOCUMENT

Constitution de l'identifiant d'un document

L'identifiant d'un document doit être inscrit dans un attribut **id** associé à l'élément racine du schéma de données. Il est constitué, lors de la création d'une instance de document, d'une racine choisie dans une liste prédéfinie et d'un code généré automatiquement par le programme. Le créateur du document peut modifier le code généré s'il souhaite assurer une identification personnalisée. Il est prévenu en cas de saisie d'un identifiant existant.

La racine est constituée d'un code producteur et d'un numéro de fonds. Ces deux éléments sont attribués par le DOSI qui en assure l'unicité au niveau national. L'unicité du code, lorsque celui-ci est modifié par l'utilisateur, doit être gérée localement.

La liste des racines prédéfinies est stockée dans un fichier externe accessible par un administrateur. Cette liste contient pour chaque racine un libellé explicatif présenté lors de la création d'un document.

Importation de documents

Les documents XML candidats à l'importation devront obligatoirement posséder un attribut **id** pour l'élément racine. Les documents importés en *ascii délimité* devront présenter l'identifiant en première

position de la liste des éléments à intégrer.

Importation/Exportation

L'application doit permettre d'importer des documents existants (format XML ou ASCII délimité) et d'exporter des dossiers complets en tenant compte des liens avec des entités externes (référence à des objets multimédia, référence à des dossiers ou parties de dossiers), et en appliquant éventuellement une transformation par l'intermédiaire d'une feuille de style (XSL/XSLT).

Importation

Des règles de transformation simples et paramétrables peuvent être éventuellement appliquées pour l'importation. Ces règles et les paramètres nécessaires à l'application de ces règles sont stockés dans un fichier externe. Les différents types de règles et leur syntaxe sont consignées dans un document en annexe du cahier des charges (chapitre *Règles standards pour l'importation normale de documents*).

Exportation

Un interface permettra de sélectionner un corpus de dossiers à partir des mêmes critères de recherches utilisés lors de la saisie (voir chapitre *Dynamique de la saisie - Mode de rappel*), de lister ce corpus et de lancer l'exportation. Des critères supplémentaires de sélection pourront être ajoutés pour tenir des valeurs d'attribut figurant dans les liens (exemple : limiter l'exportation d'un corpus aux documents dont le mode de diffusion des objets reliés est « public »).

Traitement des objets multimédia

Les objets multimédia associées au documents seront pris en compte dans le traitement de l'exportation :

- recopie de la ressource au même endroit que les fichiers XML
- possibilité de transformation pour certains types d'objets (ex : réduction de la taille des images au format d'affichage plein écran par défaut 768 x 512, changement de format pour une publication sur le WEB).

Un fichier de récolement sera généré pour permettre d'exploiter les fonds exportés sur un serveur centra-

lisé au DOSI. Ce fichier établira la correspondance entre les dossier exportés et les objets multimédia associés (référence du dossier, identifiant de l'objet).

Situation du noyau par rapport aux futurs développements

Le noyau logiciel, objet du présent cahier des charges, fera l'objet de développements complémentaires (phase n° 2) :

- gestion de la publication en ligne des dossiers validés avec système d'indexation et de recherche , interrogation dynamique de la base;
- publication papier de qualité professionnelle des dossiers ;

Bien que ces développements ne fassent pas partie du présent cahier des charges, il convient que le prestataire prenne en compte ceux-ci afin de prévoir des possibilités d'intégration de ces futurs développements autour du noyau qu'il aura développé.

Les logiciels à fournir dans le cadre de la prestation

Les logiciels à développer par le prestataire sont les suivants :

- un générateur de schéma SGBD;
- un générateur de masque de saisie ;
- une plate-forme de production intégrant un éditeur de documents, un éditeur de liens et un système d'indexation des documents;
- un programme de reprise de l'existant.

En plus des logiciels développés, le prestataire devra fournir ou proposer les logiciels complémentaires suivants :

- un éditeur de schémas de données ;
- un outil de création de feuille de style (XSL) ;
- un outil de transformation de documents (XSLT).

Ces logiciels doivent être libres. Le prestataire justifiera son choix et en précisera les limites.

Systemes cibles

Serveur

Logiciels de bases

Le serveur de documents doit s'appuyer sur :

- **LINUX** comme système d'exploitation (Red Hat ou Mandrake) ;
- **Apache** comme serveur HTTP ;

- **MySQL** ou **Postgres** comme logiciel de gestion de base de données.
- **Cocoon** pour la transformation (XSL/XSLT) de dossiers.

S'il est nécessaire d'ajouter des logiciels complémentaires à ceux précédemment cités, ceux-ci sont obligatoirement des logiciels libres.

Le prestataire doit justifier le choix des logiciels qu'il aura retenus.

Langage de programmation

Le langage utilisé pour le développement de la plate-forme est JAVA. Le choix éventuel d'un autre langage ne peut être utilisé que pour une partie limitée du système (traitements critiques) et doit être justifié par le prestataire.

Plate-forme de développement

JBUILDER dernière version. Le choix éventuel d'une plate-forme de développement doit être justifié par le prestataire.

Clients

Les clients sont obligatoirement des navigateurs (Netscape ou IE) sous LINUX et sous WINDOWS (à partir de 95).

Aucun autre logiciel ne doit être nécessaire sur les clients, seuls des plug-ins téléchargeables à partir du serveur de documents peuvent être employés.

Performances

Dans le présent chapitre, on entend par masque de saisie tout interface ou formulaire permettant de saisir des données.

Les performances lors de la saisie devront être les suivantes :

- chargement pour la première fois d'un masque de saisie ; temps inférieur à 1 seconde ;
- par la suite, chargement du même masque de saisie vide après validation de la saisie précédente ; temps inférieur à 0,5 secondes ;
- passage d'un masque à un autre via un lien ; temps inférieur à 1 seconde.
- chargement d'un dossier complet inférieur à 3 secondes.

Ces temps sont mesurés :

- pour le serveur Pentium III 800 Mhz avec 512 Mo de mémoire vive et 20 Go de disque magnétique ;
- pour le poste client Pentium III 800 Mhz avec 128 Mo de mémoire vive et 10 Go de disque dur ;

- la liaison entre le serveur et le poste client est une liaison Ethernet à 10 Mo/s.

Ces mesures sont réalisées avec un seul utilisateur.

Des tests de charge doivent démontrer qu'avec 10 utilisateurs simultanés sur un même serveurs, les performances doivent être :

- chargement pour la première fois d'un masque de saisie ; temps inférieur à 3 secondes ;
- par la suite, chargement du même masque de saisie vide après validation de la saisie précédente ; temps inférieur à 2 secondes ;
- passage d'un masque à un autre via un lien ; temps inférieur à 3 secondes.

Reprise de l'existant

Le système initial sera fourni avec un mécanisme de reprise :

1. pour les structures de données écrites actuellement sous forme de DTD à convertir en schémas XML.
2. pour les bases de données actuelles (ASCII délimité ou fichiers XML) qui devront être converties en documents XML conformément aux schémas résultant de la conversion des DTD fournies par le Ministère; Cette reprise concerne uniquement les documents qui ne pourraient être traités dans le cadre d'une importation normale de documents (voir chapitre *Règles standards pour l'importation normale de documents*).
3. pour des documents traitement de texte (à partir de couper-coller ou de conversion de feuille de style en balise XML).

Des règles de conversion devront être appliquées pour tenir compte notamment de valeurs manquantes dans les notices (exemple : prénom absent dans un champ AUTEUR). Les règles de conversion sont fournies dans un document en annexe du cahier des charge (chapitre *Règles spéciales d'importation pour la reprise de l'existant*).

Délivrables

Le prestataire doit fournir :

un CD-ROM « **programmes** » comportant

- le code source de l'application;
- les schémas XML pour les tables de paramétrage (tables des liens et spécifications des formulaires);

un CD-ROM « **installation** » comportant

- les scripts permettant les opérations de génération;
- la documentation technique (étude et réalisation) établie suivant les normes et besoins définis par le DOSI ;
- la documentation utilisateur;

un CD-ROM « **reprise** » comportant

- les scripts permettant les opérations de reprise;
- une première application en utilisant les exemples fournis en annexe.

Formation

Le prestataire assurera les formations suivantes :

1. utilisateurs participants à la validation (environ 10 personnes)
2. administration et personnalisation du système (5 personnes).

ANNEXE AU CAHIER DES CHARGES

1.1 Tests de validation à fournir par le prestataire

Le prestataire fournira un schéma de test par direction.

Les tests devront permettre de valider toutes les fonctionnalités décrites dans le cahier des charges (saisie arborescente, saisie par formulaire, utilisation des outils d'aide à la saisie, notamment la sélection de termes dans des lexiques, contrôle d'intégrité, établissement de liens internes ou externes, etc.).

Dossier d'archives (DAF)

Un dossier d'archives, autrement dit un instrument de recherche archivistique, comporte trois modules correspondant aux trois éléments englobants dans la DTD EAD :

- une description bibliographique de l'instrument de recherche proprement dit (titre, auteur, langue, date de publication ou de révision) ;
- éventuellement une description de la page de titre de l'instrument de recherche ;
- le contenu scientifique proprement dit de l'instrument de recherche, lui-même subdivisé en :
- des informations générales relatives au fonds ou à la collection d'archives décrits globalement (intitulé, identifiant, dates extrêmes des documents, nom de l'institution responsable, provenance ; informations administratives sur les conditions d'accès aux documents, sur l'historique de la conservation, sur les modalités d'entrée du fonds dans l'institution ; informations sur la personne physique ou morale producteur du fonds ; autres informations telles que références bibliographiques utiles ; points d'accès contrôlés ; description sommaire du contenu des documents) ;
- des informations relatives à chacune des subdivisions du fonds ou de la collection d'archives (par exemple des séries), elles-mêmes divisées en d'autres éléments subordonnés, et ainsi de suite, jusqu'à des descriptions très détaillées, au niveau de chaque pièce d'archives (charte médiévale, texte administratif, carte, plan, sceau, affiche...) le cas échéant. Pour chacun de ces niveaux, on pourra trouver le même type d'informations que celles fournies au niveau global (intitulé, identifiant, dates extrêmes des documents) ainsi que des éléments d'indexation ou des points d'accès contrôlés.

Le dossier d'archives pourra comporter des liens vers des objets externes, notamment vers des images numériques des documents d'archives décrits (à tous les niveaux de description) ou vers d'autres instruments de recherche ou ouvrages relatifs aux documents. Les images numériques pourront être en très grand nombre ; parfois il pourra donc être nécessaire d'appeler par un lien une application dédiée à la consultation par feuilletage de ces images.

Dossier d'artiste (DAP)

Dossiers d'artistes

le dossier d'artiste a comme caractéristique principale de parler d'une personne vivante, qui donc évolue.

La notion de **répétabilité** des éléments est primordiale.

Notion d'information **publique et restreinte**

Possibilité de signaler dans un texte un nom d'auteur, une citation, un nom d'oeuvre, une technique, étranger (langue)

MCC/DAG/DOSI/BEI/FROLLAND/MCC-DOC

Possibilité de visualiser une image avant de l'importer.

Création automatique de l'identifiant d'éléments

le dossier d'artiste est relié par des **liens** vers les ressources externes : images, bases d'oeuvres (notamment Joconde et Vidéomuseum), base de références bibliographiques, dossiers documentaires, fichiers de recherches (établies avec Syndoc), site internet, adresse mel, et par des liens « internes » (d'éléments à éléments)

Le dossier d'artiste comprend :

Identification du responsable

Identité de l'artiste ou du collectif d'artistes

dénomination , adresse, atelier, sexe, naissance, décès, nationalité, représentants (adresse), illustration, note

Parcours de l'artiste

Formation - titre, dates, organisme, note :éléments répétables autant de fois que nécessaire
Activité idem

Prix et bourses

titre et date, organisme, adresse, contact, rôle note et divers ; éléments répétables

Caractéristiques

Groupe (possibilité de l'ajouter à un autre dossier) titre/dates/description/note

Mouvement

Domaine : liste d'autorité

mots clés liste d'autorité

Thème

Les listes d'autorité sont stockées dans des lexiques. Les Lexiques devront être initialisés avant le test.

Oeuvres ou ensemble d'oeuvres

Titre, dates, copyright, technique, dimensions, location, description, note, illustration

Collection titre, sous titre, description

Expositions personnelle et collective

Titre, dates, endroit, description, notes, commissaire,(adresse), bibliographie participants

Textes et sections

Bibliographie

BiblioML représentation XML des formats bibliographiques et autorités UNIMARC

Dossier de l'Inventaire Général

Les enquêtes des équipes régionales de l'Inventaire sur le patrimoine architectural et mobilier donnent lieu à la constitution de dossiers documentaires articulés entre eux.

Il existe 5 types de dossiers :

- le **dossier généralités** est un dossier de présentation de l'étude et de son territoire : objectifs scientifiques, objets concernés, conditions d'enquête, cadre historique et géographique ;
- les **dossiers collectifs** sont les dossiers établis sur des oeuvres réunies volontairement par les enquêteurs pour répondre aux besoins d'une étude sur une catégorie d'oeuvres particulière ;

MCC/DAG/DOSI/BEI/FROLLAND/MCC-DOC

- les **dossiers d'objet d'étude « architecture »** sont des dossiers sur des œuvres particulières, qu'il s'agisse d'œuvres isolées ou d'ensembles. Ils peuvent être décomposés et accompagnés de sous-dossiers,
- les **dossiers sommaire objets mobiliers** servent d'introduction et de présentation de l'étude des œuvres mobilières contenues dans un édifice ;
- les **dossiers d'objets d'étude « objet mobilier »** sont des dossiers sur des œuvres particulières, qu'il s'agisse d'œuvres isolées ou d'ensembles. Ils peuvent être décomposés et accompagnés de sous-dossiers,

Le contenu des dossiers et les liens que l'on peut créer entre eux sont normalisés, formalisés dans un schéma unique. Dans chaque dossier, l'information est organisée en différents chapitres : identification de l'objet d'étude (désignation et localisation), historique, description... Les dossiers comprennent également des illustrations (plans, photographies, relevés...) et des références bibliographiques (sources).

Les dossiers peuvent être vus comme un assemblage d'éléments dont certains sont réutilisables (de façon autonome ou dans d'autres dossiers). Ainsi, tous les dossiers contiennent-ils une notice signalétique rédigée conformément aux consignes *des systèmes descriptifs de l'architecture* ou *des objets mobiliers*, qui seront versés dans les bases de données documentaires Mérimée (architecture) ou Palissy (objets mobiliers). Les illustrations de toute nature font l'objet d'un référencement dans la base de données locale Illustration (gestion de la photothèque), par une notice rédigée conformément *au système descriptif de l'illustration*, qui pourra être versée, accompagnée de l'image numérique correspondante dans la base nationale d'images Mémoire. Les références bibliographiques sont extraites de bases de données locales topo-bibliographiques. Enfin, les dossiers font l'objet d'un géo-référencement (intégration des coordonnées Lambert des œuvres sous forme ponctuelle, de tracé ou d'emprise).

Au moment de la constitution du dossier, les différents éléments proviennent donc de sources variées : ils peuvent être saisis directement lors de la constitution du dossier ou récupérés par import s'il s'agit d'images, de textes, d'enregistrement de bases de données (y compris géographiques) existantes. Réciproquement, les éléments saisis directement peuvent également être amenés à alimenter des bases de données.

Enfin, les dossiers contiennent des relations internes et peuvent être liés à d'autres documents électroniques.

L'objectif est de saisir un ensemble de dossiers issus des enquêtes de l'Inventaire, représentatif des 5 types de dossier existants.

La base de travail sera la DTD 2.5 de l'Inventaire (à transformer en schéma XML). Cette DTD contient la description :

- des cinq types de dossiers énumérés ci-avant (un schéma par dossier),
- des relations possibles entre chaque dossiers (un schéma pour les liens).

Par ailleurs, deux bases de données « génériques » devront être implémentées dans le système :

- base de données Illustration (schéma conforme au *système descriptif de l'illustration*), pour référencer les illustrations,
- base de données topo-bibliographique (schéma *texto2*) pour les sources bibliographiques.

Chaque dossier comprend les éléments suivants :

- une fiche signalétique comportant les éléments d'indexation du dossier (mots-clés pris dans des listes contrôlées) et une synthèse du contenu du dossier,
- plusieurs volets d'illustrations réparties par type dans différents éléments : une ou plusieurs cartes et plans, un ou plusieurs relevés, une ou plusieurs photographies... Toutes les illustrations sont référencées par une notice dans une base de données *Illustration* et réutilisables entre plusieurs dossiers, et le lien entre le dossier et l'image se fait toujours via une notice Illustration. La notice descriptive des illustrations est soit saisie au moment de la constitution du dossier, soit récupérée de ce pot commun. Le lien créé entre la notice Illustration et le dossier contient les éléments de contexte (légende).
- un volet sources bibliographiques contenant une ou plusieurs références bibliographiques, consolidées dans un pot commun réutilisable pour d'autres dossiers (saisies pour le dossier ou récupérées de ce pot commun),
- un géoréférencement,

- un ou plusieurs textes librement rédigés (introduction, synthèse, observations générales).

Le test devra permettre la saisie d'un ensemble de dossiers correspondant aux 5 types définis ci-avant et des liens que l'on peut établir entre eux : un dossier généralités d'aire d'étude, un dossier collectif (maisons-fermes), un dossier architecture (abbaye) comprenant des sous-dossiers (bâtiment conventuel et église), un dossier sommaire objets mobiliers (les objets de l'église), un dossier objets mobiliers (tombeau) comportant un sous-dossier (statue).

Les fonctions suivantes devront être testées :

- la saisie directe des différents éléments du dossier. Cette saisie se fera par l'intermédiaire de formulaires pour la fiche signalétique, les volets illustration et sources bibliographiques. Les formulaires devront inclure les aides à la saisie (liste déroulante des valeurs ou accès au thésaurus pour les éléments à vocabulaire normalisé),
- le géoréférencement des différents dossiers, par pointage sur une carte des points, tracés ou emprise des objets à localiser,
- la création de liens entre éléments du dossiers,
- la création de liens entre dossiers, avec implémentation des fonctions de recherche pour le dossier cible,
- la création de liens entre une notice Illustration et un objet multimédia,
- l'export de dossiers complets au format XML

Dossier d'œuvre

A la Direction des musées de France, actuellement trois bases de données sont concernées par le dossier électronique et traitent des oeuvres archéologiques, beaux-arts et ethnologique. Il est envisagé par la suite de construire d'autres modules spécifiques tels que le dossier restauration, le dossier musée, le dossier gestion.

1/ Présentation

Le dossier œuvre est composé de plusieurs modules, dont le module principal est la description de l'œuvre :

- module **description de l'œuvre** (numéro-inventaire, titre, désignation, technique/matière, utilisation, destination, dimensions, inscriptions et marques, sujet représenté, datation, état de l'œuvre, ancienne appartenance, textes complémentaires);
- module **artiste** (nom, prénom, surnom, école stylistique, date et lieu de naissance, date et lieu de décès);
- module **collecte** (nom géographique de site, date)
- module **bibliographique(s)**² (titre de l'ouvrage, auteur(s), éditeur, date de publication, etc.)
- module d'**exposition** (titre et date de l'exposition)
- module **image** de l'oeuvre (numéro d'inventaire de l'image, nom du CD contenant l'image numérique, nom du fichier image sur ce CD);
- module **localisation** : le musée de conservation de l'œuvre (nom de musée et adresse).
- module **statut juridique** (type de propriété, mode d'acquisition, protection, établissement acquéreur, date d'acquisition, dépôt)
- module **état** (constat d'état)

2/ Caractéristiques techniques

- En général les modules peuvent être répétables à. Il y a un ou plusieurs artistes pour une œuvre. Il y a zéro à n références bibliographiques pour une œuvre. Il y a zéro ou une notice image par œuvre. Il y a un et un seul musée de conservation pour une œuvre, mais une œuvre peut-être en dépôt.
Il peut être associé à un artiste de zéro à n références bibliographiques.
- Chaque module doit être autonome, et permettre de ne pas remplir l'ensemble des éléments de ce module si l'élément principale est déjà connu dans la base. Les liens entre les différents

modules doivent être réalisés automatiquement.

- Tous les tronçons peuvent être saisis de façon indépendante, sauf le tronçon image qui est saisi obligatoirement en dépendance du tronçon œuvre.
- Tous les champs contenant une date sont de la forme AAAA/MM/JJ.
- Le dossier d'œuvre peut contenir des liens vers des objets externes, notamment vers des images numériques, vers des adresses URL (vers des bases de données, des sites), vers des adresses électroniques (ex : commande d'image), et par des liens «internes» (un dossier vers un autre)

3/ Indexation et listes d'autorité

- Listes d'autorité
 - éléments avec des listes d'autorité propre : «matière et technique», «école stylistique» et «domaine» possèdent chacun leur liste d'autorité propre.
 - éléments avec des listes d'autorités partagées : lexique noms propres (pour les artistes, les personnes représentées, le propriétaire), lexique géographique (pour la localisation, statut juridique, collecte, etc.);
- Indexation
 - éléments indexés : «numéro-inventaire», «titre», «sujet-représenté», «nom de l'artiste», «surnom», «école stylistique», «titre de l'ouvrage», «éditeur», «date de publication», «numéro d'inventaire de l'image», «nom du CD contenant l'image numérique», «nom du musée de conservation de l'œuvre» sont des éléments indexés.
 - éléments avec du texte libre : possibilité d'identifier dans le contenu des informations relevant de lexiques existants (lexiques nom propre, géographique, matière et technique, etc.) ou de préciser la signification du contenu (titre d'une œuvre, d'une exposition, d'une publication, etc.).

4/ Les fonctions à tester

- La saisie
Faire une saisie des différents éléments du dossier par l'intermédiaire de formulaires avec des volets spécifiques à chaque module et lié à une aide à la saisie (liste déroulante des valeurs ou accès au thésaurus pour les éléments à vocabulaire normalisé).
- Les liens
Vérifier la création des liens entre les différents modules, les liens extérieurs (vers d'autres bases de données, URL, etc.)
- L'export
L'export de dossiers avec l'ensemble des modules au format XML et au format Mistral.

1.2 Méthode d'identification des dossiers et des fragments

Identification d'un dossier (à faire valider par le comité multimédia, en tenant compte de l'existant dans les bases documentaires et pour la gestion des images)

L'identifiant est constitué de 3 éléments : IDPRODNUMF_NUMP

MCC/DAG/DOSI/BEI/FROLLAND/MCC-DOC

radical

IDPROD

Identifiant de producteur du dossier. Attribué par le DOSI (numéro séquentiel, chaîne alphanumérique de longueur variable).

NUMF

Identifiant de fonds sur 4 caractères obligatoirement (4 caractères alphanumériques , par défaut 0000), défini par le producteur qui doit gérer en interne son unicité.

Code local

NUMP

Numéro de série du dossier à l'intérieur du fonds, généré par le système (15 caractères maximum , modifiable).

Identification d'un bloc à l'intérieur d'un dossier

id : numéro séquentiel de bloc à l'intérieur du dossier (sur 5 caractères) généré automatiquement.

La concaténation des codes d'identification de document et de bloc constituent une adresse de fragment unique à l'intérieur du Ministère. Exemple :

00240435_012.xml#id(A0034)

1.3 Description des contrôles

Les contrôles permettent de vérifier l'intégrité d'un document par rapport à un système descriptif. Ces contrôles mettent en évidence trois types d'erreurs :

- les erreurs syntaxiques,
- les erreurs de vocabulaire
- les erreurs de cohérence

Les règles de contrôles sont constituées d'expressions booléennes intégrant les opérateurs ET/OU/SAUF et les tests suivants :

présence/absence d'éléments
présence/absence de chaînes de caractères dans un élément ou un attribut
présence/absence de terme dans un lexique
interdiction de terme dans un élément ou un attribut
appartenance d'un terme à une famille dans un lexique
nombre de descripteurs
syntaxe (type, longueur, position)
Longueur minimum/maximum d'élément/attribut
fonctions (règles faisant appel à plusieurs contrôles simultanés)

Il est possible de faire des regroupements d'expressions (un seul niveau de parenthèses).

Chaque règle de contrôle est définie par :

MCC/DAG/DOSI/BEI/FROLLAND/MCC-DOC

- un niveau d'application (document, fragment, élément ou attribut),
- un nom d'élément ou d'attribut (dépendant du niveau d'application),
- une expression ou fonction avec arguments,
- un message en cas d'erreur,
- un domaine d'application (les fonds documentaire sont parfois découpés en domaines)
- un niveau de gravité (grave ou non grave).

Contrôle lexical

Un contrôle lexical de base consiste à vérifier que le ou les termes contenus dans un élément sont présents dans le lexique associé à cet élément. Des règles de contrôles plus sophistiquées utilisant la hiérarchie des thesauri, peuvent être mises en œuvre comme la vérification d'appartenance d'un terme à une famille de termes.

Implémentation actuelle

Les lexiques sont stockés actuellement dans une base Pervasive sur un serveur NT. L'application de contrôle OVIDE (poste client Windows) utilise une API écrite en C++ pour accéder aux thesauri en mode partagé.

Les contrôles sont écrit en C++. Ces contrôles sont paramétrables (chaînes de caractères et noms de champs variables). Un fichier d'initialisation (type INI windows) fait le lien entre les règles de contrôles et les éléments à contrôler dans les notices documentaires.

Implémentation future

Les contrôles actuels seront modélisés sous forme de schémas XML. Les fichiers de paramétrage des contrôles (.INI) existants seront convertis en instances de schémas de contrôle dans le cadre de la reprise de l'existant.

Les règles de contrôles actuelles qui peuvent être intégrées directement dans les schémas de données des documents ne seront pas prises en compte dans la modélisation (exemple : cardinalités, types). Le prestataire précisera les modalités d'intégration de ces règles dans les schémas.

1.4 Modélisation des liens

Une fusion des différents modèles de lien utilisés dans les schéma actuellement en vigueur au Ministère devra être effectuée pour permettre une utilisation générique de la plate-forme de production. Le mécanisme de liens implémenté devra permettre de mettre en œuvre des relations simples (1 à 1) comme c'est le cas actuellement pour les dossiers d'Inventaire et les dossiers d'artistes, ou étendues (relations groupées) comme il est permis de le faire avec le schéma EAD des Archives.

Exemple d'implémentation actuelle des liens dans les dossiers d'Inventaire (Sous-Sirection de l'Inventaire Général)

```
<!ELEMENT ILLUSTR (VIGNETTE)? >
<!ATTLIST ILLUSTR
  xml:link CDATA #FIXED «simple»
  show CDATA #FIXED «new»
  actuate CDATA #FIXED «user»
  href CDATA #REQUIRED
  ID ID #IMPLIED
  DESTLIEN IDREFS #IMPLIED>
```

```

<!ELEMENT VIGNETTE EMPTY >
<!ATTLIST VIGNETTE
  xml:link          CDATA          #FIXED «simple»
  show              CDATA          #FIXED «embed»
  actuate           CDATA          #FIXED «auto»
  href              CDATA          #REQUIRED >

```

Exemple d'implémentation actuelle des liens dans les dossiers d'artiste (Direction des musées de France)

```

<!ELEMENT Image EMPTY>
<!ATTLIST Image
  xRef              CDATA          #IMPLIED
  xRefSysteme      CDATA          #IMPLIED
  id                ID            #IMPLIED
  lang              NMTOKEN       #IMPLIED
  Diffusion         (Publique|Restreinte)
  Facettes          IDREFS        #IMPLIED
  Entite            ENTITY         #IMPLIED
  Source            CDATA          #IMPLIED
  MIMEType          CDATA          #IMPLIED
  Largeur           NMTOKEN       #IMPLIED
  Hauteur           NMTOKEN       #IMPLIED
  Unites            NMTOKEN       #IMPLIED
  DPI               NMTOKEN       #IMPLIED >

```

Exemple de modélisation des liens multiples dans la DTD EAD (Archives)

Exemple d'implémentation actuelle des liens simples dans la DTD EAD (Archives)

```

<!ELEMENT dao      (daodesc?)>
<ATTLIST dao
  id                ID            #IMPLIED
  altrender         CDATA          #IMPLIED
  audience          (external|internal) #IMPLIED
  entityref         ENTITY         #IMPLIED
  xpointer          CDATA          #IMPLIED
  xlink:form        CDATA          #FIXED          «simple»
  href              CDATA          #IMPLIED
  role              CDATA          #IMPLIED
  title             CDATA          #IMPLIED
  show              (embed|replace|new) #IMPLIED
  actuate           (auto|user)    #IMPLIED
  behavior          CDATA          #IMPLIED
  content-role      CDATA          #IMPLIED
  content-title     CDATA          #IMPLIED
  inline            (true|false)    «true»>

```

```

<!ELEMENT bibref
(#CDATA|ptr|extptr|enph|lb|abbr|expan|edition|imprint|name|num|bibseries|ref|title|famname|persname|corpname|extref|archref)*>
<ATTLIST bibref
  id                ID            #IMPLIED
  altrender         CDATA          #IMPLIED
  audience          (external|internal) #IMPLIED
  entityref         ENTITY         #IMPLIED
  xpointer          CDATA          #IMPLIED
  xlink:form        CDATA          #FIXED          «simple»
  href              CDATA          #IMPLIED
  role              CDATA          #IMPLIED
  title             CDATA          #IMPLIED
  show              (embed|replace|new) #IMPLIED
  actuate           (auto|user)    #IMPLIED
  behavior          CDATA          #IMPLIED
  content-role      CDATA          # I M P L I E D
  content-title     CDATA          #IMPLIED
  inline            (true|false)    «true»>

```

```
encodinganalog CDATA #IMPLIED >
```

Exemple de modélisation des liens multiples dans la DTD EAD (Archives)

```
<!ELEMENT daogrp (daodesc?,(daoloc|extptrloc|extrefloc|ptrloc|refloc)*)>
<!ATTLIST daogrp
  xlink:form CDATA #FIXED «extended»
  altrender CDATA #IMPLIED
  audience (external|internal) #IMPLIED
  content-role CDATA #IMPLIED
  content-title CDATA #IMPLIED
  id ID #IMPLIED
  role CDATA #IMPLIED
  inline (true|false) #IMPLIED >

<!ELEMENT daoloc (daodesc?)>
<!ATTLIST daoloc
  xlink:form CDATA #FIXED «locator»
  actuate (auto|user) #IMPLIED
  altrender CDATA #IMPLIED
  audience (external|internal) #IMPLIED
  behavior CDATA #IMPLIED
  id ID #IMPLIED
  entityref ENTITY #IMPLIED
  href CDATA #IMPLIED
  role CDATA #IMPLIED
  show (embed|replace|new) #IMPLIED
  title CDATA #IMPLIED
  xpointer CDATA #IMPLIED >

<!ELEMENT daodesc (address|blockquote|chronlist|head|list|note|p|table)>
<!ATTLIST daodesc
  altrender CDATA #IMPLIED
  audience (external|internal) #IMPLIED
  id ID #IMPLIED >
```

Etc.

Les objets peuvent être référencés soit par des adresses (href, xref), soit par des entités (Entite, entityref), lesquelles entités font le lien avec un fichier électronique. Le deuxième mode de référence par entité sera privilégié (gestion des fichiers externes dans une table des entités).

Exemples de modélisation des liens dans la DTD Dossiers d'artistes (DAP)

L'élément **Lien** permet d'identifier la source d'un lien hypertextuel simple (selon la terminologie Xlink :

```
<!ELEMENT Lien (#PCDATA | classeInlineSansLiens)* >
<!ATTLIST Lien
  %attCommuns;
  idref IDREF #IMPLIED
  href CDATA #IMPLIED
  Role CDATA #IMPLIED
  Titre CDATA #IMPLIED
  Affichage (Nouveau|Traiter|Remplacer) #IMPLIED
  Activation (Utilisateur|Automatique) #IMPLIED >
```

L'élément **Image** fait référence à un fichier externe contenant une image pour une illustration :

```
<!ELEMENT Image EMPTY >
<!ATTLIST Image
  %attLiens;
  %attCommuns;
  Entite ENTITY #IMPLIED
  Source CDATA #IMPLIED
  MIMETYPE CDATA #IMPLIED
  Largeur NMTOKEN #IMPLIED
```

Hauteur	NMTOKEN	#IMPLIED
Unites	NMTOKEN	#IMPLIED
DPI	NMTOKEN	#IMPLIED>

Cet élément est nécessairement vide.

L'élément **Objet** permet d'intégrer dans le document des objets externes, habituellement multimédias ou interactifs. Il s'agit de la même approche utilisée pour les applications Java ou les contenus nécessitant des modules externes (plug-ins) pour les documents HTML.

```
<!ELEMENT Objet ( Parametre* ) >
<!ATTLIST Objet
  %attLiens;
  %attCommuns;
  declare      (declare)      #IMPLIED
  classid     CDATA          #IMPLIED
  codebase    CDATA          #IMPLIED
  data        CDATA          #IMPLIED
  type        CDATA          #IMPLIED
  codetype    CDATA          #IMPLIED
  archive     CDATA          #IMPLIED
  standby     CDATA          #IMPLIED
  height      CDATA          #IMPLIED
  width       CDATA          #IMPLIED>
```

Paramétrage

Un paramétrage permettra, pour chaque type de schéma, d'identifier les éléments de type lien, de définir des libellés clairs pour chaque balise (utilisation dans les écrans de mise en œuvre des liens) et de préciser les informations obligatoires.

1.5 Importation de documents

1.5.1 Règles standards pour l'importation normale de documents

Ce chapitre est scindé en deux parties, une première partie qui fait l'inventaire des règles à respecter en ce qui concerne le format en entrée (contraintes statiques identifiés par Cx) et une deuxième partie qui précise les règles à exécuter pour effectuer la transformation des données (Rx).

1.5.1.1 Contraintes

- C1. Deux formats seront acceptés en entrée, le format ASCII délimité, dont le séparateur de zone est le trait vertical, et le format XML.
- C2. Les documents en ASCII délimité doivent être stockés dans un fichier dont la première ligne contient la liste des champs. Les notices sont séparées par un retour ligne (CR/LF).
- C3. Les documents XML doivent être conformes au schéma de données cible.
- C4. Le nom des objets externes doit respecter la syntaxe normalisée (voir chapitre *Méthode d'identification des dossiers et des fragments*).

Localisation des sources de données

- C5. Les documents à importer sont stockés dans un répertoire unique, soit sous forme de fichiers individuels (documents XML) soit sous forme de fichiers de lots (documents ASCII délimité). S'il existe des liens entre documents, tous les documents concernés doivent être présents dans le répertoire d'importation. Ce

répertoire est indiqué dans un fichier paramètre. Exemple :

```
[Loca]
PathDoc= C :\tmp
```

C6. Les objets externes liés à des documents importés sont recherchés dans des répertoires fournis en paramètres dans un fichier. Exemple

```
[Loca]
PathObjet1=E ://images/sri
PathObjet2=C :\mmedia\images\pleinecran
PathObjet3=C :\mmedia\images\vignette
```

C7. Sont exclus du traitement les répertoires sur support amovible (CDR ou disquette).

1.5.1.2 Règles

Introduction

Les règles de transformation sont stockées dans un fichier de paramétrage propre à chaque schéma de données. Leur syntaxe sera mise au point par le prestataire en fonction des besoins exprimés dans le présent document. Cette syntaxe sera la plus simple possible et fera l'objet d'un document d'utilisation intégré à la documentation fournie par le prestataire. Un document d'étude portant sur une reprise similaire est disponible et pourra être remis au prestataire pour l'aider dans la mise au point des règles et de leur syntaxe.

Des règles de transformation pourront être ajoutées à la présente liste si elles s'avèrent indispensables à une importation correcte des jeux de test fournis par le Ministère.

Le fichier de paramétrage peut contenir des lignes vides (espacement des règles) ou des lignes de commentaires. (lignes commencent par //).

1.5.1.2.1 Règles liées à l'importation de notices en ASCII délimité

R1. Les règles de transformation sont stockées dans un fichier de paramètres externe portant le même nom que le schéma de données cible.

R2. Les règles de transformation de documents en ASCII délimité suivent une logique non ambiguë respectant le schéma de données cible.

R3. Une notice source alimente un ou plusieurs documents cibles. Par contre, un document cible ne peut être alimenté que par un seul document source. Pour chaque champ source, il faut déterminer le schéma cible. Pour éviter la création de documents en doubles dans un schéma cible, il faut indiquer le ou les champs qui permettent d'identifier de façon unique le document cible (pas de création de document si un document existe déjà avec le contenu des champs d'identification). Un paramétrage permet d'indiquer les liens qui seront créés entre les différents documents générés à partir d'une seule notice (nom du schéma de document principal, nom du schéma de document annexe, nom de l'élément de lien utilisé pour relier le document principal au document annexe, titre du lien, etc).

R4. En cas d'absence de règle de transformation pour un champ, le champ est recopié dans un élément portant le même nom dans le schéma cible à condition que cet élément soit unique. Si le champ n'a pas d'équivalent dans la structure cible, il est ignoré.

R5. Les documents ne satisfaisant pas aux règles de transformation (au moins une règle non satisfaite) et au schéma de données cible sont ignorés. L'utilisateur peut consulter la liste des documents en erreur en

MCC/DAG/DOSI/BEI/FROLLAND/MCC-DOC

fin d'importation.

Identification d'un document

R6. Une règle permet de spécifier la manière dont le code identifiant d'un document est généré. Cette règle précise le mode de génération (manuel ou automatique). Si le mode de génération est manuel, la règle doit indiquer le ou les champs à prendre en compte pour la constitution du code. Dans le cas où ce code est constitué du contenu de plusieurs champs, une syntaxe permet d'en préciser la mise en forme (possibilité d'utiliser des caractères séparateurs). Cet identifiant peut alimenter le contenu d'un élément ou un attribut du document créé (nom, type et élément englobant à préciser dans la règle).

Règle de transformation d'un champ

R7. Par défaut, un élément cible est rempli avec le contenu du champ portant le même nom dans la notice.

R8. Pour chaque champ traité, un élément englobant peut être précisé dans le cas où une ambiguïté serait possible.

R9. Un élément peut recevoir le contenu d'un champ ne portant pas le même nom.

R10. Un champ peut alimenter plusieurs éléments ou attributs cible. Dans ce cas, il faut préciser pour chaque élément ou attribut cible, l'élément englobant et le nom de l'élément cible (cas où la cible est un attribut).

R11. Si un élément de sortie ne peut être satisfait par une valeur en entrée alors qu'il est obligatoire dans le schéma cible, il est possible d'indiquer une valeur par défaut.

R12. Lorsque la cible est un attribut et non un élément, le paramétrage doit permettre de l'indiquer (exemple nom_elt/nom_att). Le traitement d'un attribut suppose que l'élément associé soit rempli également, à moins que la déclaration de ce dernier dans le schéma de données précise qu'il puisse être vide (valeur nulle ou valeur constante définie dans les règles de transformation).

R13. Une règle peut être accompagnée d'une condition (présence ou absence d'un autre champ, présence ou absence d'une chaîne de caractères dans le champ traité ou dans un autre champ, utilisation de troncatures). La condition peut être une expression booléenne du type CONTIENT « a » OU NE CONTIENT PAS « b » ET CONTIENT « c » etc.

R14. Un élément peut recevoir une valeur constante en attribut en fonction d'une condition. Cet élément peut être rempli ou non (élément cible vide) par le contenu du champ indiqué.

R15. Il est possible d'empêcher le remplissage d'un élément par un champ de même nom (par exemple, l'élément est traité par une autre règle ou ne peut être traité dans le cadre de la reprise).

R16. Un élément peut recevoir le contenu de plusieurs champs. La règle permet de préciser la syntaxe à utiliser pour la concaténation des différents champs

Règle de décision

R17. En plus des règles classiques de transformation de champs en éléments, il faut pouvoir définir des règles de génération automatique d'éléments intermédiaires permettant d'interconnecter des fragments de document lorsque cette génération ne peut être effectuée normalement. Par exemple, le schéma de données de l'Inventaire possède un sous-ensemble alternatifs d'éléments sous la racine. Le programme d'importation doit être capable de déterminer lequel de ces éléments doit être pris en compte pour la constitution du document cible.

Décomposition d'un champ en structure arborescente

R18. La décomposition d'un champ multivalué en plusieurs éléments est définie par un ensemble de paramètres.

Un premier paramètre donne le nom de l'élément englobant dans le schéma de données cible (si une ambiguïté est possible, il faut donner un chemin relatif composé de plusieurs éléments afin de lever l'ambiguïté).

Le deuxième paramètre est la liste des séparateurs à utiliser pour la décomposition du champ en éléments.

Les paramètres suivants décrivent la décomposition. Pour chaque phase de la décomposition, il faut indiquer un critère de sélection, une condition d'exécution (voir règle R13) et le nom de l'élément ou de l'attribut de sortie.

Le critère de sélection peut s'exprimer de plusieurs manières :

- soit en indiquant un rang,

exemple :

prendre le terme à la position 2 pour le mettre dans l'élément PRENOM

- soit en donnant un couple de caractères séparateurs encadrant l'item,

exemple :

prendre le terme encadré par des parenthèses pour le mettre dans l'élément PRECISION

- soit en indiquant une chaîne de caractères constante à rechercher avec ou sans troncatures,

exemple :

prendre le terme qui contient la chaîne « étud »

ou un type de chaîne

exemple :

prendre le premier terme et le mettre dans l'élément COTE s'il est numérique.

autre exemple :

si le terme situé à l'intérieur de parenthèses est un chiffre, renseigner l'attribut NOMBRE, si c'est un ? renseigner l'attribut DOUTE, sinon, renseigner l'attribut COMPL.

- soit en indiquant un lexique,

exemple :

si le terme est présent dans le lexique AUTEUR, renseigner l'élément AUTEUR.

Certains critères de sélection pourront être accompagnés d'une liste de caractères séparateurs si nécessaire (découpe du champ en descripteurs et application de la règle à chaque descripteur).

R19. Lorsqu'un critère de sélection est exprimé sous forme de chaîne de caractères, il est possible d'utiliser la troncature pour indiquer que la chaîne commence ou termine un mot (ex *etud**), est exactement un mot (découpage selon les espaces et les caractères de ponctuation) ou se trouve n'importe où dans l'item (troncatures gauche et droites).

R20. Si les items en entrée alimentent un seul type d'élément en sortie, le critère de sélection n'est pas nécessaire.

R21. Il est possible de déclarer certains items « optionnels » dans une suite d'items identifiés par un rang. Le nombre d'items déclarés comme tel ne doit pas prêter à confusion (les notices traitées doivent contenir tous les items ou aucun des items optionnels).

Traitement des structures à occurrences multiples

R22. Plusieurs niveaux de décomposition peuvent être appliqués pour permettre la génération de structures à occurrences multiples à condition que cette génération ne soit pas incompatible avec le schéma de données cible. Dans ce cas, une liste de séparateurs et un sous-ensemble d'éléments cibles doit être indiqué pour chaque décomposition.

Exemples

Champ **AUTR**

Pour un champ **AUTR** qui peut contenir un ou plusieurs auteurs séparés par des point-virgules, chaque auteur étant décrit par son nom, suivi d'une virgule, suivie du prénom et d'une période entre parenthèses indiquée par deux dates Début et Fin.

Exemple : *AUTR:DUPONT, Jean (1949-1978) ; DURAND (2001-)*

Une première décomposition est appliquée avec le séparateur point-virgule. Chaque item obtenu est décomposé avec les séparateurs virgule, tiret et parenthèses.

Les items optionnels **Prenom** et **Fin** étant susceptibles d'être absents séparément, une erreur d'interprétation risque de se produire si l'on définit tous les items en entrée uniquement par leur rang. En conséquences, les deux premiers items sont identifiés par une position, les items **Debut** et **Fin** étant identifiés par un couple de caractères d'encadrement (« parenthèse ouvrante, tiret » pour Debut et « tiret, parenthèse fermante » pour Fin).

Une autre possibilité consiste à éliminer la virgule et le tiret du deuxième niveau et à spécifier un troisième niveau de décomposition avec d'une part la virgule et d'autre part le tiret.

L'élément **Pays** est défini par une constante car il ne figure pas systématiquement dans le champ.

Résultat :

```
<Auteurs>
  <Auteur>
    <Nom>DUPONT</Nom>
    <Prenom>Jean</Prenom>
    </Periode Debut=»1949" Fin=»1978">
    <Pays>France</Pays>
  </Auteur>
  <Auteur>
    <Nom>DURAND</Nom>
    </Periode Debut=»2001">
    <Pays>France</Pays>
  </Auteur>
</Auteurs>
```

champ **MATR** :

exemple : *MATR :laine(?),soie(chaine,trame):satin,brodé ;laiton(garniture):repoussé*

Les techniques satin et brodé sont mises en facteur avec les matériaux laine et soie.

L'élément englobant **Matériaux** peut contenir plusieurs matériaux. Chaque élément **Matériau** peut avoir plusieurs caractéristiques. Les éléments **Technique** et **Doute** ne sont pas obligatoires. L'élément **Groupe** permet de conserver la mise en facteur des techniques pour un groupe de matériaux.

Résultat :

```
<Matériaux>
  <Groupe>
    <Matériau doute=»oui»><Nom>laine</Nom></Matériau>
    <Matériau>
      <Nom>soie</Nom>
      <caractéristique>chaine</caractéristique>
      <caractéristique>trame</caractéristique>
    </Matériau>
    <Technique>satin</Technique>
    <Technique>brodé</Technique>
  </Groupe>
  <Groupe>
    <Matériau>
      <Nom>laiton</Nom>
      <caractéristique>garniture</caractéristique>
      <Technique>repoussé</Technique>
    </Matériau>
  </Groupe>
</Matériaux>
```

Le point d'interrogation indiquant une incertitude sur le terme soie est remplacé dans la structure cible par l'attribut doute avec la valeur prédéfinie « oui ».

champ **MICR** :

Ce champ est composé d'une succession de segments (séparateur de segment le point-virgule). Chaque segment est introduit par le terme « microfiche » puis entre parenthèses une information décomposée en trois parties : le type d'étude, le numéro de microfiche dans la commune, les coordonnées de la première page (pas de séparateur pour ces informations).

exemples : *MICR :microfiche (PI 1/23 B 12) ; microfiche (2/3 D 4)*

```
<MICROFICHE>
  <MICR>
    <TYPETUD >PI</ TYPETUD >
    <NUM>1/23 </NUM >
    <COOR>B 12 </COOR >
  </MICR>
  <MICR>
    <NUM>2/3 </NUM >
    <COOR>D 4 </COOR >
  </MICR>
</MICROFICHE>
```

La règle est la décomposition du champ en éléments MICR selon le séparateur point virgule. Chaque segment (on ne garde que le contenu de la parenthèse) est ensuite décomposé en deux ou trois éléments :

- si le premier descripteur commence par un chiffre : décomposition en deux éléments (premier descripteur dans l'élément NUM, ce qu'il y a après le blanc dans l'élément COOR)
- sinon : premier descripteur dans l'élément TYPETUD, le second dans l'élément NUM et la suite dans l'élément COOR.

N.B : Toutes ces règles constituent un cadre général de transformation. Les documents qui ne satisferaient ces règles devront être ramenées au cas général par une transformation en amont.

1.5.1.2.2 Génération des liens vers des objets externes

Reconnaissance des champs source des notices en ASCII délimité

- R23. Un paramétrage permet pour chaque lien de définir la correspondance entre les champs source et les éléments cible **href**, **titre**, **diffusion** normalisés (voir chapitre *modélisation des liens*). Le contenu de l'élément **rôle**, et le séparateur de valeurs si le champ est multivalué.
- R24. L'élément **href** qui précise le nom de l'objet externe est rempli avec le contenu du champ indiqué en paramètre.
- R25. Une valeur par défaut peut être précisée pour les éléments **titre** et **diffusion**.
- R26. Le champ **diffusion** est rempli avec le contenu du champ indiqué en paramètre, à condition que ce champ contienne une valeur cohérente avec le schéma des données (publique ou restreinte). Si ce n'est pas le cas, le lien n'est pas créé et l'erreur est signalée dans un fichier journal.
- R27. Si le champ alimentant **href** est multivalué, autant de liens sont créés que de valeurs trouvées après découpage selon le séparateur indiqué dans le paramètre **sep**. Les informations titre et diffusion sont les mêmes pour tous les liens créés.
- R28. Plusieurs profils de lien peuvent être définis à condition qu'il y ait une seule occurrence du couple href/champ.

Exemple de paramètres : *cas d'un lien entre un document et une image*

```
[Lien]
href=VIDEO
titre=LEGEND|néant
role=image
diffusion=DIFF|Restreinte
sep=;
```

L'élément href du nouveau lien contient le nom de la ressource image contenu dans le champ VIDEO de la notice. Si le champ VIDEO contient deux noms d'images séparés par un point-

virgule, 2 liens sont créés avec le même titre extrait du champ LEGEND et le même mode de diffusion.

Si l'objet est un document, deux cas peuvent se présenter. Soit le document est un fichier texte externe et dans ce cas, le traitement est le même que pour un objet multimédia de type image. Soit le document est une notice incluse dans l'un des lots à importer. Le système devra, dans ce cas, créer un lien entre les deux documents importés.

Cas d'un lien à établir entre deux notices importés

```
[Lien]
id=REFA
titre=|édifice
role=document
diffusion=DIFF|Publique
```

Le lien est établi entre deux documents créés à partir de deux notices importées. La référence de la notice cible est indiquée dans le champ REFA de la notice source. L'élément **id** du nouveau lien contient l'identifiant du document créé à partir de la notice cible. Le champ REFA est obligatoirement de type mono-valeur. Le titre du lien est fixé à la valeur « édifice ».

R29. En cas de non résolution d'un lien, les documents concernés ne sont pas importés .

1.5.1.2.3 Règles liées à l'importation de documents XML

R30. L'éclatement d'un seul documents XML en entrée vers plusieurs schémas de données en sortie est exclue de l'importation normale (voir chapitre *reprise de l'existant*).

1.5.2 Règles spéciales d'importation pour la reprise de l'existant

Les documents qui ne peuvent être traités normalement par le module d'importation seront traités par un programme spécial de reprise de l'existant. Le présent chapitre établit la liste des points qui devront être traités lors de cette reprise. Les règles précisées dans le précédent paragraphes s'appliquent à la reprise de l'existant également dans la mesure où elles ne sont pas incompatibles avec les spécifications de conversion énoncées dans le présent chapitre.

Traitement des documents XML contenant plusieurs schémas de données

Un algorithme permettra de générer plusieurs documents en sortie à partir d'un seul document en entrée. Par exemple, si la plate-forme gère séparément les schémas pour un dossier d'œuvre et pour un dossiers d'artiste et que les documents en entrée regroupent en un seul document la vie d'un artiste et toutes ses œuvres, plusieurs documents devront être générés, un document pour l'artiste et autant de documents que d'œuvres décrites. Les liens seront générés automatiquement entre les différents documents produits (liens du dossier d'artiste vers les dossiers d'œuvres). Les schémas de données dans les documents source et cible peuvent être différents. Les règles définies pour l'importation normale de documents seront utilisables également pour le passage de fragment de données d'un schéma à l'autre.

1.5.2.1 Cas spécifique de reprise des notices des Archives

Le CHAN produit aujourd'hui deux types de documents électroniques concernés par le présent CCTP :

- des instruments de recherche archivistiques réalisés, soit en XML en conformité avec l'EAD, soit, plus souvent, sous Word, sans recourir à une base de données. Ces derniers seront transformés par le CHAN en documents XML bien formés et valides, conformes à la DTD EAD avant d'être importés dans la

plate-forme décrite dans le présent CCTP. Aucune règle de reprise de l'existant n'est donc nécessaire pour ces documents. La plate-forme permettra ensuite de les produire directement en XML.

- des bases de données ou banques d'images. Il en existe aujourd'hui une quinzaine susceptibles d'être gérées et consultées en ligne. Ces bases ont actuellement des structures et des contenus hétérogènes, qui seront transformés et normalisés pour aboutir à un seul modèle de données. Les règles de reprise de l'existant définies ci-dessous concernent les données en ASCII délimité conformes à ce modèle. Par commodité, dans la suite de ce chapitre, on appellera ce modèle base de données.

Le programme de reprise de l'existant recevra en entrée plusieurs fichiers ASCII délimité, qu'il traitera séparément pour générer un document XML conforme à l'EAD pour chacun de ces fichiers.

Chacun des fichiers ASCII contiendra un nombre variable de notices exportées de la base de données du CHAN.

On trouvera ci-dessous la liste des champs de la base de données et, le cas échéant, le nom des éléments cible dans le document XML :

Nom du champ dans l'arborescence de l'EAD	Type de champ	Nom de l'élément ou de l'attribut cible
Remarques		1.5.2.1.1.1.1
REF	Champ clé	Attribut id de ead\archdesc\dsc\c
Le contenu du champ REF ne sera importé dans le document XML que pour les notices dont le champ NIVEAU contient une valeur de suffixe différent de (1)		
NIVEAU	Monovalué	Contient une valeur et un suffixe constitué d'un chiffre de 1 à n entre parenthèses
Idem		Attribut level de ead\archdesc OU de Ead\archdesc\dsc\c
PNIVEAU	Multivalué	
Le contenu de ce champ n'est utile que pour le traitement des données lors de la reprise. Il n'est pas nécessaire de l'importer dans le document XML.		
TITREINV	Texte	ead\eadheader\filedesc\titlestmt\titleproper
AUTRINV	Texte	ead\eadheader\filedesc\titlestmt\author
DATPBINV	AAAA	ead\eadheader\filedesc\publicationstmt\date
DATRVINV	AAAA	ead\eadheader\revisiondesc\change\date
DOMAINE	Monovalué	Ead\archdesc\did\unittitle
DATEX	Texte	Ead\archdesc\did\unitdate
COTEX	Texte	Ead\archdesc\did\unitid
OU		
Ead\archdesc\dsc\c\did\unitid		
Le choix de l'élément cible dépend de la valeur du champ NIVEAU		
DEPOT	Monovalué	Ead\archdesc\did\repository
ORIGINE	Texte	Ead\archdesc\did\origination
PHYSDESC	Texte	Ead\archdesc\did\physdesc
OU		
Ead\archdesc\dsc\c\did\physdesc		
Le choix de l'élément cible dépend de la valeur du champ NIVEAU		
ACCESSIB	Texte	Ead\archdesc\admininfo\accessrestrict\p
OU		

ead\archdesc\dsc\c\admininfo\accessrestrict\p		Idem
REPRODUC	Texte	Ead\archdesc\admininfo\userrestrict\p
OU Ead\archdesc\dsc\c\admininfo\userrestrict\p		Idem
Nom du champ	Type de champ	Nom de l'élément ou de l'attribut cible
dans l'arborescence de l'EAD		1.5.2.1.1.2
<i>Remarques</i>		
ORGANIS	Texte	Ead\archdesc\scopecontent\organization\p
BIBREF	Texte	Ead\archdesc\add\otherfindaid\p
COTCONSU	Monovalué	Ead\archdesc\dsc\c\did\unitid
COTORIGI	Monovalué	Ead\archdesc\dsc\c\did\note\p
DATDEB	AAAA-MM-JJ	Ead\archdesc\dsc\c\did\unitdate
Si les deux champs ont le même contenu, la valeur ne sera importée qu'une fois. Si les deux valeurs sont différentes, elles seront importées dans le même élément <unitdate>, la première précédée de "Entre le", la deuxième précédée de "et le"		
DATFI	AAAA-MM-JJ	Ead\archdesc\dsc\c\did\unitdate
TAILLE	Texte	Ead\archdesc\dsc\c\did\physdesc\dimensions
FORME	Multivalué	Ead\archdesc\did\physdesc\physfacet
OU		Ead\archdesc\dsc\c\did\physdesc\physfacet Valeur de l'attribut type de physfacet : «form»
Le choix de l'élément cible dépend de la valeur du champ NIVEAU		
SUPPORT	Multivalué	Ead\archdesc\did\physdesc\physfacet
OU		Ead\archdesc\dsc\c\did\physdesc\physfacet Valeur de l'attribut type de physfacet : «material»
Idem		
TECHNIQ	Multivalué	Ead\archdesc\did\physdesc\physfacet
OU		Ead\archdesc\dsc\c\did\physdesc\physfacet Valeur de l'attribut type de physfacet : «technique»
Idem		
ANALYSE	Texte	Ead\archdesc\dsc\c\scopecontent\p
NOTES	Texte	Ead\archdesc\dsc\c\scopecontent\note\p
VIDEO	Multivalué	Attribut href de Ead\archdesc\dsc\c\dao
TYPEDOC	Multivalué	Ead\archdesc\controlaccess\genreform
OU		Ead\archdesc\dsc\c\controlaccess\genreform
Le choix de l'élément cible dépend de la valeur du champ NIVEAU		
PERSPHYS	Multivalué	Ead\archdesc\controlaccess\persname
OU		Ead\archdesc\dsc\c\controlaccess\persname
Idem		
PERSMORA	Multivalué	Ead\archdesc\controlaccess\corpname
OU		Ead\archdesc\dsc\c\controlaccess\corpname
Idem		
Nom du champ	1.5.2.1.2	Type de champ
dans l'arborescence de l'EAD		Nom de l'élément ou de l'attribut cible
		1.5.2.1.2.1
<i>Remarques</i>		

FONCTIONS	Multivalué	Ead\archdesc\controlaccess\occupation
OU		Idem
LIEUX	Multivalué	Ead\archdesc\controlaccess\geogname OU
		Idem
MOT-MAT	Multivalué	Ead\archdesc\controlaccess\subject OU
		Idem
NOTHIS	Texte	Ead\archdesc\dsc\c\odd\p
NOTBIB	Texte	Ead\archdesc\dsc\c\add\otherfindaid\p

Chaque notice contient tout ou partie des champs énumérés ci-dessus.
 Dans les champs multivalués, le séparateur de valeurs est le point-virgule.

Pour chaque fichier ASCII délimité, le programme de reprise réalisera successivement les traitements suivants :

1.5.2.1.3 PHASE 1

- Recherche et sélection de la notice dont le champ NIVEAU contient en suffixe la valeur (1). Il y aura une seule fiche de ce type pour chaque fichier ASCII délimité.
- Importation du contenu de cette notice dans les éléments <eadheader> et éléments fils, <archdesc> et éléments fils suivant le tableau.
 - Dans les cas signalés par le tableau où plusieurs éléments cible sont possibles pour le contenu d'un champ, le programme choisira la première option.
 - L'attribut level de ead\archdesc sera donc rempli avec le contenu du champ NIVEAU, dont le programme retranchera le suffixe (1).
 - Pour les champs multivalués TYPEDOC, PERSPHYS, PERSMORA, FONCTIONS, LIEUX, MOT-MAT, le programme créera autant d'éléments cible de même nom qu'il y a de valeurs séparées par le point-virgule dans le champ, et importera chacune des valeurs dans un des éléments ainsi créé.
 - Pour les champs multivalués FORME, SUPPORT et TECHNIQ, le programme importera intégralement le contenu de chaque champ dans un élément ead\archdesc\did\physdesc\physfacet différent. Il y aura donc au plus trois éléments <physfacet> de contenu différent. Le programme donnera à l'attribut type de <physfacet> la valeur définie dans le tableau.

A la fin de cette phase le programme aura créé et rempli l'élément <eadheader> du document XML et tous ses éléments subordonnés, ainsi que les éléments de haut niveau (<did>, <add>, <admininfo>, <scopecontent> et éventuellement <controlaccess>) de <archdesc>.

1.5.2.1.4 PHASE 2

- Recherche et sélection des notices dont le champ NIVEAU contient en suffixe la valeur (2). Il y aura de 1 à n notices de ce type dans le fichier ASCII délimité.
- Importation du contenu de chacune de ces n notices dans un élément <c> différent de ead\archdesc\dsc et ses éléments subordonnés, suivant le tableau.
 - Il y aura donc après traitement autant d'éléments <c> créés sous ead\archdesc\dsc que de notices de ce type.

- L'attribut level de chaque élément ead\archdesc\dsc\c sera donc rempli avec le contenu du champ NIVEAU dont le programme retranchera le suffixe (2)
- Dans les cas signalés par le tableau où plusieurs éléments cible sont possibles pour le contenu d'un champ, le programme choisira la deuxième option
- Le contenu du champ PNIVEAU est seulement utilisé pour la reprise de l'existant
- Pour les champs multivalués TYPEDOC, PERSPHYS, PERSMORA, FONCTIONS, LIEUX, MOT-MAT, le programme créera autant d'éléments cible de même nom qu'il y a de valeurs séparées par le point-virgule dans le champ, et importera chacune des valeurs dans un des éléments ainsi créé.
- Pour les champs multivalués FORME, SUPPORT et TECHNIQ, le programme importera intégralement le contenu de chaque champ dans un élément ead\archdesc\dsc\c\physdesc\physfacet différent. Il y aura donc au plus trois éléments <physfacet> de contenu différent. Le programme donnera à l'attribut type de <physfacet> la valeur définie dans le tableau.
- Pour les champs DATDEB et DATFI, deux cas pourront se présenter :
 - soit les deux champs contiennent la même valeur ; dans ce cas cette valeur commune sera importée une seule fois dans Ead\archdesc\dsc\c\did\unitdate ;
 - soit chacun des champs contient une valeur différente ; dans ce cas les deux valeurs seront importées dans le même élément <unitdate>, la valeur du champ DATDEB en premier, précédée de la chaîne de caractères «Entre le» , la valeur du champ DATFI en second, précédée de la chaîne de caractères «et le». Ex. : <unitdate> Entre le 1418-01-01 et le 1418-12-31</unitdate>
- Pour le champ multivalué VIDEO, qui contient le nom et le chemin d'accès des images numériques, deux cas peuvent se présenter :
 - soit VIDEO contient au plus 6 valeurs différentes, qui désignent, pour trois documents source différents, la vignette et l'image plein écran
Le programme créera donc six éléments <dao> différents.
Le premier élément <dao> permettra d'atteindre la première vignette indiquée dans le champ VIDEO (première valeur du champ de suffixe _V.EXT) ; l'attribut href de cet élément sera rempli avec cette valeur ; le deuxième élément <dao> permettra d'atteindre l'image plein écran correspondant à la vignette (autre valeur du champ, de même préfixe que la vignette mais de suffixe différent) ; etc
Pour chaque élément <dao> correspondant à une vignette, le programme donnera à l'attribut show la valeur «embed», à l'attribut actuate la valeur «auto», à l'attribut xlink:form la valeur «simple» et à l'attribut title la valeur «vignette».
 - soit VIDEO contient plus de 6 valeurs différentes
Dans ce cas qui pourra être fréquent, le programme écrira dans le document XML les informations nécessaires pour appeler une application de visualisation par feuilletage des images numériques désignées dans VIDEO.

A la fin de cette phase le programme aura créé et rempli les éléments <c> correspondant à un deuxième niveau de description archivistique.

S'il n'y a pas dans le fichier ASCII d'autres notices que celles déjà importées, le programme fermera les éléments <c> créés, les éléments <dsc> et <archdesc> englobant ces éléments <c>, puis l'élément <ead> racine et enregistrera le document XML. Le traitement sera alors achevé.

1.5.2.1.4.1 EXEMPLE

Ci-dessous, on trouvera un exemple illustrant ces deux phases et présentant :

A : un fichier au format ajout piloté de TEXTTO, contenant 4 notices, 1 de niveau 1 et 3 de niveau 2

B : le document XML à produire à partir de ce fichier.

A : fichier au format ajout piloté de TEXTO (pour plus de clarté, on a choisi ce format dans l'exemple ; il est entendu que les fichiers à traiter seront au format ASCII délimité)

REF . SCEAUX
NIVEAU . collection (1)
PNIVEAU sceaux de Bourgogne
DOMAINE . collection des sceaux de Bourgogne
TITREINV. Inventaire analytique de la collection des sceaux de Bourgogne
AUTRINV . Inventaire préparé par le SNTIC du CHAN
DATPBINV. 2001
DATRVINV. 2001
DATEX . 0847-1740 (circa)
COTEX . SC/B1 à SC/B1610
DEPOT . Centre historique des Archives nationales
ORIGINE . Collection des moulages de sceaux de Bourgogne réalisés par les Archives nationales... (etc)
PHYSDESC. 2 000 moulages de plâtre pris sur les empreintes originales
BIBREF . Les sceaux décrits ici ont été inventoriés en 1912 par Coulon dans «Inventaire des sceaux de la Bourgogne, recueillis dans les dépôts d'archives, musées et collections particulières des départements de la Côte-d'Or, de Saône-et-Loire et de l'Yonne» publié à Paris, réimprimé par Kraus-Thomson en 1980
ACCESSIB. moulages consultables en salle de lecture du CARAN
REPRODUC. reproduction des moulages possible par le service des sceaux ; photographies ou impressions d'images numériques
ORGANIS . La collection se divise en sceaux laïques et ecclésiastiques, eux-mêmes répartis en séries : souverains, grands dignitaires de la couronne, ducs, comtes et vicomtes, grands dignitaires (Bourgogne, Champagne), seigneurs, villes et bourgeois, métiers, cours et juridictions, offices, cardinaux et clergé romain, archevêques et évêques, chapitres, dignités et offices capitulaires, paroisses, universités et collèges, abbayes, prieurés, ordres religieux, ordres religieux militaires, maladreries et confréries, divers et inconnus.
TYPEDOC document en trois dimensions ; moulage de sceau ; sceau
MOT-MAT . sigillographie

REF . SCEAUX02
NIVEAU . item (2)
PNIVEAU . sceaux de Bourgogne ; sceau
COTCONSU. SC/B1
DATDEB . 0847-08-25
DATFI . 0847-08-25
FORME . ovale
TAILLE . hauteur 40 mm x largeur 32 mm
ANALYSE . Moulage du sceau de Charles II le Chauve, roi de France.
Légende
en latin restituée : «KAROLUS GRATIA DEI REX», traduction :
Charles, roi par la grâce de Dieu.
TYPEDOC . sceau de souverain
PERSPHYS. Charles II le Chauve (roi de France)
LIEUX FRANCE - IXe siècle
MOT-MAT . Moyen Age ; France sous les Carolingiens ; sceau en buste

; personnage en buste ; homme ; profil ; couronne ; vêtement à l'antique ; vêtement masculin

FONCTIONS roi de France

NOTES . Le document portant le sceau original est conservé aux Archives départementales de Saône-et-Loire.

REF SCEAUX03

NIVEAU item (2)

PNIVEAU sceaux de Bourgogne ; sceau

COTCONSU SC/B58

DATDEB 1285-04-02

DATFI 1285-04-02

FORME rond

TAILLE diamètre 70 mm

ANALYSE Moulage du sceau d'Hugues de Bourgogne, seigneur de Montréal, fils d'Hugues IV, duc de Bourgogne. Légende abîmée et restituée en latin : «SIGILLUM HUGONIS DE BURGUNIDA DOMINI MONTIS REGALIS», traduction : sceau d'Hugues de Bourgogne, seigneur de Montréal.

TYPEDOC sceau de seigneur

PERSPHYS BOURGOGNE, Hugues de (seigneur de Montréal) - fils d'Hugues IV (duc de Bourgogne)

LIEUX MONTREAL, SEIGNEURIE DE (duché de Bourgogne) - XIIIe siècle ; MONTREAL (Yonne) ; YONNE (Bourgogne - département) ; BOURGOGNE (France - région administrative) ; France ; BOURGOGNE (France - duché) - XIIIe siècle

MOT-MAT Moyen Age ; France sous les Capétiens ; héraldique ; sceau équestre de guerre ; homme ; personnage à cheval ; cheval ; armement ; bouclier armorié

NOTES Le document portant le sceau original est conservé aux Archives Départementales de Côte-d'Or.

NOTHIS Hugues de Bourgogne (1260-1288) épouse en 1279 Marguerite de Chalon (? - 1328). Il est le fils d'Hugues IV de Bourgogne et de Béatrix de Champagne ; il est quelquefois appelé Huguenin.

REF SCEAUX04

NIVEAU item (2)

PNIVEAU sceaux de Bourgogne ; sceau

COTCONSU SC/B1562

DATDEB 1418-01-01

DATFI 1418-12-31

FORME rond

TAILLE diamètre 40 mm

ANALYSE Moulage du sceau du chapitre de l'Hôpital de Saint-Jean de Jérusalem en Champagne. Légende abîmée et restituée en latin : «SIGILLUM HOSPITALIS PRIORATUS CAMPANIE», traduction : sceau du prieuré de l'Hôpital en Champagne.

TYPEDOC sceau ecclésiastique

PERSMORA CHAPITRE DE L'HOPITAL DE SAINT-JEAN DE JERUSALEM EN CHAMPAGNE ; ORDRE DES HOSPITALIERS DE SAINT-JEAN DE JERUSALEM

LIEUX CHAMPAGNE (France - comté) - XVe siècle ; FRANCE

MOT-MAT Moyen Age ; France sous les Valois ; Xve siècle ; héraldique ; sceau emblématique ; sceau armorial ; écu armorié ; armes de Champagne ; animal ; oiseau ; aigle ; aigle de saint Jean ; prieuré ; chapitre

NOTES Le document portant le sceau original est conservé aux Archives Départementales de la Côte-d'Or.

NOTHIS Il s'agit peut-être de Reims où se trouvait le chef-lieu de la commanderie pour la Champagne et le comté d'Auxerre.

1.5.2.1.4.1.1 B : document XML à produire

Nota : XXXX.xml est le nom du document XML créé, tel que défini page 9 de la présente annexe.

```
<ead id=»XXXX.xml»>
<eadheader>
  <eadid>XXXX.xml</eadid>
  <filedesc>
    <titlestmt>
      <titleproper>Inventaire analytique de la collection des sceaux de Bourgogne</titleproper>
      <author>Inventaire préparé par le SNTIC du CHAN </author>
    </titlestmt>
    <publicationstmt>
      <date> 2001</date>
    </publicationstmt>
  </filedesc>
  <revisiondesc>
    <change>
      <date> 2001 </date>
    </change>
  </revisiondesc>
</eadheader>

<archdesc level=»collection»>
  <did>
    <unittitle> collection des sceaux de Bourgogne </unittitle>
    <unitdate> 0847-1740 (circa) </unitdate>
    <unitid> SC/B1 à 1610 </unitid>
    <repository>Centre historique des Archives nationales</repository>
    <origination> Collection des moulages de sceaux de Bourgogne réalisés par les Archives
nationales ... (etc) </origination>
    <physdesc> 2 000 moulages de plâtre pris sur les empreintes originales. </physdesc>
  </did>
  <add>
    <otherfindaid>
      <p> Les sceaux décrits ici ont été inventoriés en 1912 par Coulon dans «Inventaire des
sceaux de la Bourgogne, recueillis dans les dépôts d'archives, musées et collections particulières des départe-
tements de la Côte-d'Or, de Saône-et-Loire et de l'Yonne» publié à Paris, réimprimé par Kraus-Thomson en
1980. </p>
    </otherfindaid>
  </add>
  <admininfo>
    <accessrestrict><p>moulages consultables en salle de lecture du CARAN </p></accessrestrict>
    <userrestrict><p>reproduction des moulages possible par le service des sceaux ; photogra-
phies ou impressions d'images numériques </p></userrestrict>
  </admininfo>
  <scopecontent>
    <organization>
      <p>La collection se divise en sceaux laïques et ecclésiastiques, eux mêmes répartis
en séries : souverains, grands dignitaires de la couronne, ducs, comtes et vicomtes, grands
dignitaires (Bourgogne, Champagne), seigneurs, villes et bourgeois, métiers, cours et juri-
dictions, offices, cardinaux et clergé romain, archevêques et évêques, chapitres, dignités et
```

offices capitulaires, paroisses, universités et collèges, abbayes, prieurés, ordres religieux, ordres religieux militaires, maladreries et confréries, divers et inconnus. </p>

</organization>

<controlaccess>

<genreform> document en trois dimensions </genreform>

<genreform> sceau </genreform>

<genreform> moulage de sceau </genreform>

<subject> sigillographie </subject>

</controlaccess>

<dsc type="combined">

<c level="item" id="SCEAUX02">

<did>

<unitid>SC/B1</unitid>

<unitdate>0847-08-25</unitdate>

<physdesc><physfacet type="form">ovale</physfacet>

<dimensions>hauteur 40 mm x largeur 32 mm </dimensions>

</physdesc>

</did>

<scopecontent>

<p>Moulage du sceau de Charles II le Chauve, roi de France. Légende en latin restituée : "KAROLUS GRATIA DEI REX", traduction : Charles, roi par la grâce de Dieu.</p>

<note><p>Le document portant le sceau original est conservé aux Archives départementales de Saône-et-Loire</p></note>

</scopecontent>

<controlaccess>

<genreform>sceau de souverain</genreform>

<persname>Charles II le Chauve (roi de France)</persname>

<geogname>France – IXe siècle </geogname>

<subject>Moyen Age</subject>

<subject>France sous les Carolingiens</subject>

<subject>sceau en buste</subject>

<subject>personnage en buste</subject>

<subject>homme</subject>

<subject>profil</subject>

<subject>couronne</subject>

<subject>vêtement à l'antique</subject>

<subject>vêtement masculin</subject>

<occupation>roi de France</occupation>

</controlaccess>

</c>

<c level="item" id="SCEAUX03">

<did>

<unitid>SC/B58</unitid>

<unitdate>1285-04-02</unitdate>

<physdesc><physfacet type="form">rond</physfacet>

<dimensions>diamètre 70 mm</dimensions>

</physdesc>

</did>

<scopecontent>

<p>Moulage du sceau d'Hugues de Bourgogne, seigneur de Montréal, fils d'Hugues IV, duc de Bourgogne. Légende abîmée et retsituée en latin : "SIGILLUM HUGONIS DE BURGUNDIA DOMINI MONTIS REGALIS", traduction : sceau d'Hugues de Bourgogne, seigneur de Montréal.</p>

```

        <note><p>Le document portant le sceau original est conservé aux Archives dé-
partementales de Côte-d'Or</p></note>
    </scopecontent>
    <odd>
        <p>Hugues de Bourgogne (1260-1288) épouse en 1279 Marguerite de Chalon
(? – 1328). Il est le fils d'Hugues IV de Bourgogne et de Béatrix de Champagne ; il est quelquefois appelé
Huguenin.</p>
    </odd>
    <controlaccess>
        <genreform>sceau de seigneur</genreform>
        <persname>BOURGOGNE, Hugues de (seigneur de Montréal) – fils d'Hugues
IV (duc de Bourgogne)</persname>
        <geogname>MONTREAL, SEIGNEURIE DE (duché de Bourgogne) – XIIIe siè-
cle </geogname>
        <geogname>MONTREAL (Yonne) </geogname>
        <geogname>YONNE (Bourgogne – département)</geogname>
        <geogname>BOURGOGNE (France – région administrative)</geogname>
        <geogname>France</geogname>
        <geogname>BOURGOGNE (France – duché) – XIIIe siècle</geogname>
        <subject>Moyen Age</subject>
        <subject>France sous les Capétiens</subject>
        <subject>héraldique</subject>
        <subject>sceau équestre de guerre</subject>
        <subject>homme</subject>
        <subject>personnage à cheval</subject>
        <subject>cheval</subject>
        <subject>armement</subject>
        <subject>bouclier armorié</subject>
    </controlaccess>
</c>

    <c level=»item» id=»SCEAUX04">
        .....etc....
</archdesc>
</ead>

```

1.5.2.1.5 PHASE 3

Si le programme trouve dans le fichier ASCII n notices dont le champ NIVEAU contient le suffixe (3), il poursuivra le traitement comme indiqué ci-dessous :

- sélection des n notices dont le champ NIVEAU contient en suffixe la valeur (3), qu'on appellera pour la suite notices de niveau 3

- tri des notices de niveau 3 en fonction de la valeur de position 2 dans le champ PNIVEAU. Pour l'ensemble des notices de niveau 3, il y a autant de valeurs de position 2 différentes dans le champ PNIVEAU qu'il y avait auparavant de notices de niveau 2, donc qu'il y a dans le fichier XML d'éléments c.

- pour un lot de notices de niveau 3 partageant la même valeur V2 de position 2 dans le champ PNIVEAU, le programme cherchera dans le document XML l'élément <c> contenant les données importées de la notice de niveau 2 correspondante. Cette notice de niveau 2 correspondante est celle dont le champ PNIVEAU contient en position finale la valeur V2.

- le programme créera dans cet élément <c> autant d'éléments <c> subordonnés qu'il y a de notices dans le lot, et importera dans ces éléments cibles <c>, et dans leurs éléments subordonnés, les données contenues dans les notices du lot, suivant le tableau. Cette opération sera répétée pour chaque lot de notices.

- Il y aura donc après traitement autant de nouveaux éléments <c> créés sous ead\archdesc\dsc\c que de notices de niveau 3.
- L'attribut level de chaque élément ead\archdesc\dsc\c\c sera donc rempli avec le contenu du champ NIVEAU dont le programme retranchera le suffixe (3)
- Dans les cas signalés par le tableau où plusieurs éléments cible sont possibles pour le contenu d'un champ, le programme choisira la deuxième option
- Le contenu du champ PNIVEAU est seulement utilisé pour la reprise de l'existant
- Pour les champs multivalués TYPEDOC, PERSPHYS, PERSMORA, LIEUX, MOT-MAT, le programme créera autant d'éléments cible de même nom qu'il y a de valeurs séparées par le point-virgule dans le champ, et importera chacune des valeurs dans un des éléments ainsi créé.
- Pour les champs multivalués FORME, SUPPORT et TECHNIQ, le programme importera intégralement le contenu de chaque champ dans un élément ead\archdesc\dsc\c\physdesc\physfacet différent. Il y aura donc au plus trois éléments <physfacet> de contenu différent. Le programme donnera à l'attribut type de <physfacet> la valeur définie dans le tableau.
- Pour les champs DATDEB et DATFI, deux cas pourront se présenter :
 - soit les deux champs contiennent la même valeur ; dans ce cas cette valeur commune sera importée une seule fois dans Ead\archdesc\dsc\c\did\unitdate ;
 - soit chacun des champs contient une valeur différente ; dans ce cas les deux valeurs seront importées dans le même élément <unitdate>, la valeur du champ DATDEB en premier, précédée de la chaîne de caractères «Entre le» , la valeur du champ DATFI en second, précédée de la chaîne de caractères «et le». Ex. : <unitdate> Entre le 1418-01-01 et le 1418-12-31</unitdate>
- Pour le champ multivalué VIDEO, qui contient le nom et le chemin d'accès des images numériques, deux cas peuvent se présenter :
 - soit VIDEO contient au plus 6 valeurs différentes, qui désignent, pour trois documents source différents, la vignette et l'image plein écran
Le programme créera donc six éléments <dao> différents.
Le premier élément <dao> permettra d'atteindre la première vignette indiquée dans le champ VIDEO (première valeur du champ de suffixe _V.EXT) ; l'attribut href de cet élément sera rempli avec cette valeur) ; le deuxième élément <dao> permettra d'atteindre l'image plein écran correspondant à la vignette (autre valeur du champ, de même préfixe que la vignette mais de suffixe différent) ; etc
Pour chaque élément <dao> correspondant à une vignette, le programme donnera à l'attribut show la valeur «embed», à l'attribut actuate la valeur «auto», à l'attribut xlink:form la valeur «simple» et à l'attribut title la valeur «vignette».
 - soit VIDEO contient plus de 6 valeurs différentes
Pour chaque élément <dao> correspondant à une image plein écran, le programme donnera à l'attribut show la valeur «new», à l'attribut actuate la valeur «user», à l'attribut xlink:form la valeur «simple» et à l'attribut title la valeur «image plein écran».
- soit VIDEO contient plus de 6 valeurs différentes
Dans ce cas qui pourra être fréquent, le programme écrira dans le document XML les informations nécessaires pour appeler une application de visualisation par feuilletage des images numériques désignées dans VIDEO.

A la fin de cette phase le programme aura créé et rempli les éléments <c> correspondant à un troisième niveau de description archivistique.

S'il n'y a pas dans le fichier ASCII d'autres notices que celles déjà importées, le programme fermera les éléments <c> créés, les éléments <dsc> et <archdesc> englobant ces éléments <c>, puis l'élément <ead> racine et enregistrera le document XML. Le traitement sera alors achevé.

1.5.2.1.6 PHASE 4 et suivantes :

Le programme répètera la même suite d'opérations pour la phase 4, afin d'importer les notices dont le

champ NIVEAU contient un suffixe (4), etc.

Le critère de tri des notices de niveau n (ayant dans le champ NIVEAU un suffixe de valeur (n)) sera la valeur V de position n – 1 dans le champ PNIVEAU. Pour chaque lot de notices partageant la même valeur Vn-1 de position n-1 dans le champ PNIVEAU, le programme cherchera dans le document XML l'élément <c> contenant les données importées de la notice de niveau n-1 correspondante. Cette notice de niveau n-1 correspondante est celle dont le champ PNIVEAU contient en position finale la valeur Vn-1.

C'est dans cet élément <c> que le programme créera autant d'éléments subordonnés <c> pour y importer le contenu des notices du lot.

Cette opération sera répétée jusqu'à ce que le programme ne trouve plus de notices à importer dans le fichier. Le programme fermera alors tous les éléments encore ouverts et enregistrera le document XML.

Des exemples de fichier source et de document XML résultat pourront être fournis pour illustrer les règles relatives aux phases 3 et suivantes.

1.5.2.2 Cas spécifique de reprise des dossiers de la DAP

La Direction des arts plastiques a déjà produit un certain nombre de dossiers électroniques au format XML qui devront être intégrés à la plate-forme de production. Ces dossiers ont la particularité d'utiliser différents types de liens qu'il conviendra de prendre en compte lors de l'importation de ces dossiers dans le cadre de la reprise de l'existant.

Quatre types de liens sont susceptibles d'être rencontrés :

1. des liens de types «hypertextes», avec un élément **Lien**. Celui-ci «pointe» vers :

- le ID d'un autre élément du même fichier, ou d'un autre fichier.
- ou une URL

Exemple :

```
<Para>Participation au séminaire de l'OFAJ et aux
  <Lien idref=>exp048">
    Rencontres Internationales de la photographie
    «Le portrait comme paysage...»
  </Lien>, à Arles et à Bonn.
</Para>
...
<Exposition id=>exp048" Diffusion=>Publique» Type=>Collective»>
  <Titre>Le portrait comme paysage</Titre>
  <Date>1982</Date>
  <Endroit>
    <Nom>Galerie Arena</Nom>
    <Adresse>
      <Ville>Arles</Ville>
      <Pays Code=>fr»>France</Pays>
    </Adresse>
  </Endroit>
</Exposition>
```

1. des liens d'élément à élément avec certains attributs créés à cet effet. Par exemple, l'élément **Participant** à un attribut **Dossier**, qui, une fois renseigné avec l'attribut ID du dossier d'artiste correspondant, lie l'élément Participant au Dossier d'artiste le concernant. (Idem pour txtArtiste,...).

2. des liens vers des «images» avec l'élément **Image**. La référence au fichier peut se faire :

- avec une entité, et dans ce cas il faut utiliser l'attribut Entite.
- avec un nom de fichier, en fait un URI, et dans ce cas il faut utiliser l'attribut Source. Ce sera le cas le plus

fréquent.

Exemple :

```
<Illustration id=»ill001" Diffusion=»Publique»>
  <Titre>Sans titre</Titre>
  <Copyright><Para>Photographies de Laurent Septier</Para>
</Copyright>
  <Image Source=»00001_img_003_v.jpg» Largeur=»70" Hauteur=»100"
  Role=»Vignette»/>
  <Image Source=»00001_img_003_e.jpg» Largeur=»211" Hauteur=»300"
  Role=»Ecran»/>
</Illustration>
```

4. des liens vers des «objets» avec l'éléments **Objet**. L'élément **Objet** permet d'intégrer dans le document des objets externes, habituellement multimédias ou interactifs. Il s'agit de la même approche utilisée pour les applications Java ou les contenus nécessitant des modules externes (plug-ins) pour les documents HTML.

1.5.2.3 Cas spécifique de reprise des notices de l'Inventaire

Le schéma des données de l'Inventaire (DTD Ci 2.5) possède une structure générique qui s'applique à plusieurs types de documents. Un branchement est effectué au plus haut niveau de la hiérarchie des éléments pour choisir entre les type de documents suivants :

- Classeur d'aire d'étude topographique
- Classeur d'aire d'étude thématique
- Dossier collectif
- Dossier d'architecture
- Dossier sommaire objets
- Dossier d'objet mobilier.

Le programme de reprise doit être en mesure, d'après les valeurs contenues dans certains champs des notices à importer, d'en déduire la branche à choisir dans l'une de celles décrites plus haut pour pouvoir construire le dossier.

Le programme devra être paramétré de manière à permettre une modification des règles de gestion sans intervention dans le code. Les paramètres seront stockés dans un fichier externe.

Pour la constitution d'un même dossier, plusieurs documents sources (notices) peuvent être assemblés, issus de plusieurs bases de données (exemple : une notices « architecture » et n notices d'illustration).

Règles de reprise demandées par La sous-direction de l'Inventaire

La sous-direction de l'Inventaire possède principalement trois bases de données qu'elle souhaite reprendre dans l'outil de production de dossiers électroniques. Ces trois bases sont MERIMÉE, PALISSY et ILLUSTRATION. Pour chacune de ces bases, le système descriptif se scinde en quatre groupes de champs, les champs monovaleur, les champs multivaleur, les champs multi-séparateurs et les champs à texte intégral.

Les champs à texte intégral seront copiés tel quel dans un élément³ du schéma cible.

Les champs monovaleur possèdent un seul terme qui alimente naturellement l'élément ou l'attribut correspondant dans le schéma cible.

Les champs multivaleurs (sans affixe) doivent être découpés selon un séparateur prédéfini et rangés dans des éléments identiques sous un même terme générique dans le schéma cible.

Les champs multi-séparateurs (et les champs multivaleurs avec affixes) doivent être découpés en tenant compte d'une liste de séparateurs prédéfinis et rangés dans un fragment du schéma cible.

la base Palissy

La base Palissy contient 69 champs.

- **15 sont des champs monovalués :**

REF, DENQ, REFE, REG, DPT, COM, INSEE, AIRE, CANT, REFA, ZONE, IMPL, INTE, DMIS, DMAJ,

- **17 sont des champs texte libre :**

TITR, APPL, TICO, PAPP, LIEU, ADRS, EDIF, EMPL, VOLLS, DESC, PREP, PDIM, PETA, PINS, HIST, PPRO, OBS,

- **12 sont des champs multivaleurs (sans affixe) :**

RENV, COPY, DBOR, NOMS, DOSS, PDEN, NART, REFP, CATE, DATE, PROT, MFICH

le principe de transformation sera de définir un élément à occurrences multiples, englobé sous un élément générique :

NOMS rédacteur1 ; rédacteur2 ; rédacteur3

deviendrait

```
<NOMS>
  <NOM>rédacteur1</NOM>
  <NOM>rédacteur2</NOM>
  <NOM>rédacteur3</NOM>
</NOMS>
```

- **6 sont des champs multivaleurs avec affixe ? ou numérique :** DENO, PART, PARN, ETAT, SCLE, STAT

l'affixe ? ou numérique doit être considéré comme un attribut de l'information (doute sur l'information, complément pour la caractériser en nombre). La proposition serait de raisonner comme pour le cas précédent, en ajoutant en plus deux attributs à l'élément décomposé.

Le champ DENO : croix de sacristie (?) ; chasubles (3)

deviendrait un élément DENOMINATION décomposé en autant d'éléments DENO qu'il y a de termes, chaque élément DENO pouvant être affecté des attributs DOUTE et NOMBRE

```
<DENOMINATION>
  <DENO DOUTE = « ? »>croix de sacristie</ DENO >
  <DENO NOMBRE = « 3 »>chasubles</ DENO >
</ DENOMINATION >
```

La règle de transformation doit permettre de mettre le contenu de la parenthèse en attribut et de choisir l'attribut cible selon la valeur de la parenthèse (? dans DOUTE, sinon dans NOMBRE).

NB : il existe des cas particuliers (rares), avec deux affixes - *DENO : mesures à grains (? , 2)*.

- **19 sont des champs multi-séparateurs, à traiter au cas par cas**

ETUD, PLOC, COOR, COORM, DEPL, STRU, MATR, REPR, DIMS, INSC, AUTR, AFIG, ATEL, PERS, EXEC, ORIG, STAD, DPRO, MICR

champ ETUD

Ce champ est composé de séquences (séparateur point-virgules) constituées de deux parties : un type d'étude (liste fermée) et un affixe (liste ouverte). La proposition serait la structure cible suivante :

+

exemple

ETUD : inventaire topographique ; enquête thématique régionale (monuments aux morts de la Somme)

```
<ETUDES>
  <ETUD>
    <TYPETUD>inventaire topographique</ TYPETUD >
  </ ETUD >
  <ETUD>
    < TYPETUD >thématique régionale</ TYPETUD >
    <LIBELLE>monuments aux morts de la Somme </LIBELLE>
  </ ETUD >
</ETUDES>
```

Remarque : dans la DTD CI 2.5, ce champ est un simple attribut de la notice de dossier généralités (ce qui ne permet pas de raffinement dans la structure, et donc pas de contrôle).

champ PLOC

Ce champ est composé de deux parties (séparateur point-virgules) la première concerne le type de précision (liste fermée), la seconde est l'explicitation de la précision : un terme introducteur (liste fermée) et un libellé. La structure est la suivante :

+, TYPE

+

PLOC : commune fusionnée après inventaire ; anciennement commune de : Opoul

```
<PLOC TYPE = «commune fusionnée après inventaire»>
  <INTRO> anciennement commune de </INTRO>
  <LIBELLE> Opoul </LIBELLE>
</PLOC>
```

La règle est donc : premier segment du champ (avant le ;) dans l'attribut TYPE et le second segment réparti entre l'élément INTRO (avant le :) et l'élément LIBELLE (après le :).

NB : pour Mérimée, il existe une notion complémentaire : « œuvre située sur plusieurs communes ». Il faut indiquer cette information dans l'attribut TYPE et mettre l'indication du nom de la commune dans l'élément LIBELLE.

champ COOR

ce champ est composé de deux parties (séparateur point-virgules) la première est l'ordonnée et la seconde l'abscisse. La structure est la suivante :

COOR : 0438600 ; 1200950

```
<COOR>  
  <X> 0438600</X>  
  <Y> 1200950</Y>  
</COOR>
```

la règle serait donc : premier segment du champ (avant le ;) dans l'élément X et le second dans l'élément Y.

champ COORM

Ce champ est composé d'une succession de couples de coordonnées. Le séparateur entre les couples est le /, le séparateurs à l'intérieur d'un couple est le ; La structure est la suivante :

+ , *TYPE*

+

COORM : 0401500 ; 3334500/0401500 ; 3335200/0402320 ; 3335200/0402320 ; 3334500/0401500 ; 3334500

```
<ETENDUE TYPE =» POLYGONE «>  
  <COORM>  
    <X>0401500</X>  
    <Y>3334500</Y>  
  </COORM>  
  <COORM>  
    <X>0401500 </X>  
    <Y>3335200</Y>  
  </COORM>  
  <COORM>  
    <X>0402320</X>  
    <Y>3335200</Y>  
  </COORM>  
  <COORM>  
    <X>0402320</X>  
    <Y>3334500</Y>  
  </COORM>  
  <COORM>  
    <X>0401500</X>  
    <Y>3334500</Y>  
  </COORM>  
</ETENDUE>>
```

la règle serait

- si le premier couple est le même que le dernier : l'attribut TYPE vaut «POLYGONE», sinon «POLYLIGNE»,

- chaque couple (séparés par des /) donne lieu à la création d'un élément COORM, le premier segment du couple (avant le ;) générant un élément X et le second un élément Y.

champs DEPL, EXEC, ORIG (+ REMP de Mérimée)

Refonte dans une structure commune des champs DEPL, EXEC, ORIG (+ **REMP de Mérimée**) dans une même structure.

1- Pour traiter les œuvres volées

(cas où la notice contient le descripteur « œuvre volée » ou « œuvre volée partiellement » dans le champ DEPL)

=> **Création d'un éléments VOL, avec un attribut TYPVOL** (contenant le mot clé issu du champ DEPL : descripteurs « œuvre volée » ou « œuvre volée partiellement »), élément à texte libre contenant le texte du champ VOLS de la grille PALISSY.

La règle de conversion consisterait à générer un élément VOL contenant le contenu du champ (texte intégral) VOLS, avec pour attribut TYPVOL le contenu de DEPL (« œuvre volée » ou « œuvre volée partiellement »).

exemple :

DEPL : œuvre volée partiellement

VOLS : un des trois pleurants, exposé dans le vestibule, fut volé en 1992.

<VOL TYPE= « œuvre volée partiellement » >

un des trois pleurants, exposé dans le vestibule, fut volé en 1992

</VOL>

2. création d'une structure générique pour DEPL, ORIG, EXEC (+REMP de Mérimée)

+, TYPMVT, PREC, DOUTE

HIST

L'élément MOUVEMENTS contiendrait autant éléments mouvements que de segments (séparateurs ;) dans les champs EXEC, ORIG et DEPL (et pour Mérimée, REMP).

Chaque élément MVT serait affecté :

- d'un attribut sur le type de mouvement TYPMVT, dont la valeur dépendra du champ d'où provient l'information : « lieu d'exécution » (EXEC), « lieu de déplacement » (DEPL), « lieu de destination » ou « lieu de provenance » (en fonction du terme introducteur dans le champ ORIG)...

- d'un attribut PREC pour une éventuelle précision « partie en emploi ».

- d'un attribut de doute

- de la décomposition du lieu en éléments unitaires dans l'élément LIEU (attribut du LIEU : HIST pour préciser s'il s'agit d'un « nom historique »)

exemple :

DEPL lieu de déplacement : Alsace, 67, Wissembourg, mairie

EXEC nom historique du lieu d'exécution : Pays d'Ouche (?)

ORIG lieu de provenance : Haute-Normandie, 27, Bus-Saint-Rémy, abbaye du Trésor (partie en emploi)

<MOUVEMENTS>

<MVT TYPMVT = «*lieu de déplacement*» >

<LIEU>

<PAYS> *France* </PAYS>

<REG> *Alsace*</REG>

<DPT> *67* </DPT>

<COM> *Wissembourg* </COM>

<EDIF> *mairie* </EDIF>

</LIEU>

</MVT>

<MVT DOUTE= «*?*» TYPMVT «*lieu d'exécution*» >

<LIEU HIST= «*nom historique*» >

<COM> *Pays d'Ouche* </COM>

</LIEU>

</MVT>

<MVT TYPMVT = «*lieu de provenance*» PREC = «*partie en emploi*» >

<LIEU>

<PAYS> *France* </PAYS>

<REG> *Haute-Normandie* </REG>

<DPT> *27* </DPT>

<COM> *Bus-Saint-Rémy* </COM>

<EDIF> *abbaye du Trésor* </EDIF>

</LIEU>

</MVT>

</MOUVEMENTS>

A noter sur la décomposition du LIEU : l'exemple : *pays d'Ouche* ne rentre pas dans la décomposition REG DPT COM . Peut-on le mettre directement dans l'élément ?

champ STRU

Le contenu de ce champ varie en fonction du domaine traité (vitrail, mobilier, patrimoine industriel...).

Ce champ est composé d'une succession de descripteurs :

+ , *NOMBRE* , *DOUTE*

+

exemple

STRU : jours de réseau (5) ; lancettes (en plein-cintre) ; panneaux (6)

<STRUCTURE>

champ REPR

Ce champ est composé d'une succession de segments relatifs à des supports (séparateurs la livre £). Pour chaque support, la description de la représentation se fait dans le cadre de « phrases » (séparateur de phrase : le point-virgule ;) correspondant à une signification. La phrase est composée de mots-clés, agencés selon une syntaxe faisant intervenir les séparateurs parenthèses (), virgule et deux-points :

+

+

+

exemple

REPR : incendie : église, guerre de 1914-1918, village, fuite £ Assomption, les apôtres £ Nativité, Calvaire, Ascension ; I.H.S. ; scène (cycle narratif : dévotion, paysan, statue, village, source)

```
< REPRESENT >
  <SUPPORT>
    <REPR>
      <MOTCLE> incendie </MOTCLE>
      <MOTCLE> église </MOTCLE>
      <MOTCLE> guerre de 1914-1918 </MOTCLE>
      <MOTCLE> village </MOTCLE>
      <MOTCLE> fuite </MOTCLE>
    </REPR>
  </SUPPORT>
  <SUPPORT>
    <REPR>
      <MOTCLE> Assomption </MOTCLE>
      <MOTCLE> les apôtres </MOTCLE>
    </REPR>
  </SUPPORT>
  <SUPPORT>
    <REPR>
      <MOTCLE> Nativité </MOTCLE>
      <MOTCLE> Calvaire, </MOTCLE>
      <MOTCLE> Ascension </MOTCLE>
    </REPR>
  <REPR>
    <MOTCLE> I.H.S. </MOTCLE>
  </REPR>
  <REPR>
    <MOTCLE> scène </MOTCLE>
    <MOTCLE> cycle narratif </MOTCLE>
    <MOTCLE> dévotion </MOTCLE>
    <MOTCLE> paysan </MOTCLE>
    <MOTCLE> statue </MOTCLE>
    <MOTCLE> village </MOTCLE>
    <MOTCLE> source </MOTCLE>
```

```

        </REPR>
    </SUPPORT>
</REPRESENT >

```

La règle est la décomposition du champ en autant d'éléments SUPPORT qu'il y a de segments séparés par des £
 Pour chacun de ces segments, on crée autant d'éléments REPR qu'il y a de « phrases » séparées par des ;
 Enfin, chaque descripteur (séparateur parenthèses, virgules et deux-points) génère un élément REPR.

champ DIMS

Ce champ est à traiter conjointement avec le champ PDIM (texte libre). La proposition serait de les regrouper sous un même élément englobant DIMENSION, répétable pour chaque partie dont on veut donner les dimensions (e.g. : un élément DIMENSION pour la statue, un autre pour le tableau...).

Le champ DIMS est composé d'une succession de segments (séparateurs entre segments, le point-virgule) relatifs aux différentes dimensions de l'objet étudié. Chaque dimension est un binôme composée d'un TYPE (=> attribut) : longueur, largeur... et d'une valeur.

+

+, *TYPDIM*

exemple

PDIM : dimensions de la statue
DIMS : h = 134 ; la = 45 ; pr = 18

```

< DIMENSION >
  <PDIM> dimensions de la statue </PDIM>
  <DIMS TYPDIM= «hauteur»>
    134
  </DIMS>
  <DIMS TYPDIM = «largeur»>
    45
  </DIMS>
  <DIMS TYPDIM = «profondeur»>
    18
  </DIMS>
</ DIMENSION >

```

La règle est la décomposition du champ en autant d'éléments DIMENSION qu'il y a de segments séparés par des ; Pour chaque segment, le premier descripteur correspond à la valeur de l'attribut *TYPDIM* (la = largeur etc), le second descripteur (après le signe égal) est la valeur de l'élément.

champ INSC

Ce champ serait à traiter conjointement avec le champ PINS, en fusionnant les deux au sein d'un même élément INSCRIPTION

Ce champ est composé d'une succession de descripteurs :

+

+

exemple

INSC : signature (peinte, sur l'oeuvre) ; inscription concernant l'iconographie (partiellement illisible) ; poinçon de maître (non identifié)

```
<INSCRIPTION>
  <INSC>
    <TYPINSC> signature</ TYPINSC >
    <COMPL> peinte</COMPL>
    <COMPL> sur l'oeuvre</COMPL>
  </INSC >
  <INSC>
    <TYPINSC> inscription concernant l'iconographie </ TYPINSC >
    <COMPL> partiellement illisible </COMPL>
  </INSC >
  <INSC>
    <TYPINSC> poinçon de maître </ TYPINSC >
    <COMPL> non identifié </COMPL>
  </INSC >
</INSCRIPTION>
```

La règle est la décomposition du champ en autant d'éléments INSC qu'il y a de segments séparés par des ; Pour chaque segment, le premier descripteur est mis dans l'élément TYPINSC, les descripteurs suivants (entre parenthèse, séparés éventuellement par des virgules) mis dans l'élément COMPL.

L'intégration du champ PINS risque d'être plus complexe : on peut avoir plusieurs types d'inscriptions : comment savoir à laquelle affecter la transcription présente dans PINS ? pour la reprise, on pourrait mettre systématiquement l'élément PINS directement sous l'élément racine DIMENSION.

champ AUTR

+, *DOUTE, TYPE*

+

exemple

AUTR : Cousin, Jean, le fils (? , peintre) ; Gellée, Claude, dit : Le Lorrain (peintre) ; Fantin-Latour, Victoria, née : Dubourg (peintre) ; Baldung, Hans, dit : Baldung, Grien (? , école de, peintre)

```

<AUTEUR>
  <AUTR DOUTE= « ? »>
    <NOM> Cousin </NOM>
    <PRENOM> Jean </PRENOM>
    <COMPL> le fils </COMPL>
    <PROFESSION> peintre </PROFESSION >
  </AUTR>
  <AUTR>
    <NOM> Gellée </NOM>
    <PRENOM> Claude </PRENOM>
    <PSEUDO>
      <NOM> Le Lorrain </NOM>
    </PSEUDO>
    <PROFESSION> peintre </PROFESSION >
  </AUTR>
  <AUTR>
    <NOM> Fantin-Latour </NOM>
    <PRENOM> Victoria </PRENOM>
    <NOMJF> Dubourg </NOMJF>
    <PROFESSION> peintre </PROFESSION >
  </AUTR>
  <AUTR DOUTE= « ? » TYPE = « école de »>
    <NOM> Baldung </NOM>
    <PRENOM> Hans </PRENOM>
    <PSEUDO>
      <NOM> Baldung </NOM>
      <PRENOM> Grien </PRENOM>
    </PSEUDO>
    <PROFESSION> peintre </PROFESSION >
  </AUTR>
</AUTEUR>

```

La règle est la décomposition du champ en autant d'éléments AUTR qu'il y a de segments séparés par des ;
 Pour chaque segment, la règle générale est la suivante : le premier descripteur est mis dans l'élément NOM, le suivant dans l'élément PRENOM, les descripteurs entre parenthèses mis dans l'attribut PROFESSION (sauf si ?, mis dans l'attribut DOUTE)

Cas particuliers à traiter :

- présence d'une particule dans le nom (dans le champ AUTR, elle se trouve en général après le prénom) : à rejeter dans l'élément PARTIC,
- présence d'un complément au nom : « fils », rang dans la hiérarchie... (présent dans le segment entre le prénom et la parenthèse de la profession) : mettre dans un élément COMPL,
- présence d'un pseudonyme (présent dans le segment entre le prénom et la parenthèse de la profession) : terme introducteur « dit » ou « dite » : créer un élément PSEUDO, décomposé lui aussi en NOM et PRENOM,
- présence d'un nom de jeune-fille : (présent dans le segment entre le prénom et la parenthèse de la profession) : terme introducteur « , née », créer un élément NOMJF,
- présence dans la parenthèse des termes « école de » ou « atelier de » : à mettre dans l'attribut TYPE.

champ AFIG : même syntaxe que pour AUTR

Ce champ est composé d'une succession de descripteurs :

+, DOUTE, TYPE

exemple

AFIG : Michel Claude, dit : Clodion (d'après, sculpteur)

```
<AUTEUR SOURCE>
  <AUTR TYPE = «d'après»>
    <NOM> Michel </NOM>
    <PRENOM>Claude</PRENOM>
    <PSEUDO>
      <NOM>Clodion</NOM>
    </PSEUDO>
    <PROFESSION>sculpteur</PROFESSION >
  </AUTR>
</AUTEUR SOURCE>
```

La règle est la même que pour le champ AUTR.

Particularité : dans la parenthèse, le premier descripteur correspond à l'attribut TYPE (« d'après »), le second à la profession. Il peut bien sûr y avoir doute.

Il y aurait bien sûr intérêt à fusionner les descriptions des champs AUTR et AFIG, la distinction pouvant se faire par le TYPE affecté à l'élément AUTR (« d'après »)

champ ATEL

Ce champ est composé d'une succession de segments (séparateurs entre segments, le point-virgule) relatifs aux différents lieux de production de l'objet étudié. Chaque lieu de production est un binôme composée d'un TYPE et d'un NOM. Le doute est permis.

+, *DOUTE, TYPATEL*

exemple

ATEL : Gobelins (?, manufacture)

```
< ATELIER >
  <ATEL DOUTE= « ?» TYPATEL= «manufacture»>
    Gobelins
  </ATEL>
< ATELIER >
```

La règle est la décomposition du champ en autant d'éléments ATEL qu'il y a de segments séparés par des ;

champ PERS

Ce champ est composé d'une succession de descripteurs :

+, *DOUTE*

exemple

PERS : Lorraine, Claude de, duc de Guise (commanditaire) ; Girardon, François (?, propriétaire)

```
<PERSONNALITE>
  <PERS>
    <NOM> Lorraine </NOM>
    <PARTIC> de </PARTIC>
    <PRENOM> Claude </PRENOM>
    <TITRE> duc de Guise </TITRE>
    <QUALITE> commanditaire </QUALITE>
  </PERS>
  <PERS DOUTE=» ?» >
    <NOM> Girardon </NOM>
    <PRENOM> François </PRENOM>
    <QUALITE> propriétaire </QUALITE>
  </PERS>
</PERSONNALITE >
```

La règle est la décomposition du champ en autant d'éléments PERS qu'il y a de segments séparés par des ; Pour chaque segment, le premier descripteur est mis dans l'élément NOM, le second dans l'élément PRENOM, le troisième (si présent) dans l'élément PARTIC, le quatrième (si présent) dans l'élément TITRE. Les descripteurs entre parenthèses sont mis dans l'élément QUALITE (sauf le point d'interrogation : attribut DOUTE de l'élément PERS).

Proposition pour la suite : traiter dans la même structure AUTR, AFIG et PERS (mettre un TYPE distinctif pour caractériser le lien à l'œuvre, fusionnant les types rencontrés sur les trois champs : « d'après », « atelier », « commanditaire », etc.).

champ STAD

Ce champ est composé de deux parties : un descripteur indiquant le stade de la création, complété éventuellement par un affixe. La proposition serait la structure cible suivante :

DOUTE

exemples :

STAD : carton (de tapisserie)

```
<STAD>
  <ETAT> carton </ETAT>
  <AFFIXE> de tapisserie </AFFIXE>
</STAD>
```

STAD : œuvre unique (?)

```
<STAD DOUTE = « ? »>
  <ETAT> œuvre unique </ETAT>
</STAD>
```

La règle est la décomposition du champ en deux éléments en fonction du séparateur (). L'affixe ? est mis dans l'attribut DOUTE s'il est présent.

NB : le contenu de ce champ paraît très hétérogène en fonction de la catégorie d'objets.

champ DPRO

Ce champ est composé d'une succession de segments (séparateur de segment le point-virgule). Chaque segment est composé de deux parties : la date correspondant à la protection, complétée par la nature de la protection (séparateur en date et nature : la virgule).

La proposition serait la structure cible suivante :

+

exemples :

DPRO : 1907/08/23 : classé au titre objet ; 1840 : classé au titre immeuble

```
<PROTECTION>
  <DPRO>
    <DATE>1907/08/23 </ DATE >
    <TYPROT> classé au titre objet </ TYPROT >
  <DPRO>
  <DPRO>
    <DATE>1840</ DATE >
    <TYPROT> classé au titre immeuble </ TYPROT >
  <DPRO>
</ PROTECTION >
```

La règle est la décomposition du champ en segments DPRO selon le séparateur point virgule. Chaque segment est ensuite décomposé en deux éléments en fonction du séparateur deux-points.

champ MICR

Ce champ est composé d'une succession de segments (séparateur de segment le point-virgule). Chaque segment est introduit par le terme « microfiche » (intérêt ? ne pas reprendre) puis entre parenthèses une information décomposée en trois parties : le type d'étude, le numéro de microfiche dans la commune, les coordonnées de la première page (pas de séparateur pour ces informations).

La proposition serait la structure cible suivante :

+

exemples :

MICR microfiche (PI 1/23 B 12) ; microfiche (2/3 D 4)

```
<MICROFICHE>  
<MICR>  
  <TYPETUD >PI</ TYPETUD >  
  <NUM>1/23 </NUM >  
  <COOR>B 12 </COOR >  
</MICR>  
<MICR>  
  <NUM>2/3 </NUM >  
  <COOR>D 4 </COOR >  
</MICR>  
</MICROFICHE>
```

La règle est la décomposition du champ en éléments MICR selon le séparateur point virgule. Chaque segment (on ne garde que le contenu de la parenthèse) est ensuite décomposé en deux ou trois éléments :

- si le premier descripteur commence par un chiffre : décomposition en deux éléments (premier descripteur dans l'élément NUM, ce qu'il y a après le blanc dans l'élément COOR)
- sinon : premier descripteur dans l'élément TYPETUD, le second dans l'élément NUM et la suite dans l'élément COOR.

la base Mérimée

La base Mérimée contient 82 champs.

- **13 sont des champs monovalués :**

REF, DENQ, REG, DPT, COM, INSEE, AIRE, CANT, REFE, ZONE, RFPA, DMIS, DMAJ,

- **15 sont des champs texte libre :**

APPL, TICO, LIEU, ADRS, EDIF, CADA, HYDR, HIST, DESC, PREP, TYPO, PPRO, OBS, PSTA, PAFF

- **33 sont des champs multivalués (sans affixe) :**

RENV, ARCHEO, COPY, DBOR, NOMS, DOSS, DENO, GENR, PDEN, VOCA, ACTU, REFP, IMPL, DATE, JDAT, JATT, PLAN, ETAG, VOUT, ELEV, COUV, ENER, VERT, ETAT, PROT, APRO, MHPP, SITE, INTE, REMA, AFFE, VISI, MFICH

- **4** sont des **champs multivaleurs avec affixe simple (? ou numérique)** :

SCLD, MURS, TOIT, STAT

- **17** sont des **champs multi-séparateurs, à traiter au cas par cas**

ETUD, PLOC, COOR, COORM, AUTR, PERS, REMP, DEPL, REPR, DIMS, DPRO, MICR sont communs avec Palissy.

COLL, PART, SCLE, ESCA, TECH sont à étudier plus précisément (cf. ci-après).

champ COLL

Ce champ est composé de trois parties : le chiffre INSEE du bâti, le nombre de repérés, le nombre d'étudiés :

exemple :

COLL : 1500 bâti INSEE ; 250 repérées ; 24 étudiées

```
<COLL>
  <INSEE> 1500 </INSEE>
  <REPERE> 250 </REPERE>
  <ETUDIE> 24 </ETUDIE >
</COLL>
```

La règle est la décomposition du champ entre trois éléments (séparateur ;).

champ PART

Ce champ contient les différentes parties constituantes (séparateur ;), avec la précision pour chacune si elle a été étudiée.

Ce champ est composé de trois parties : le chiffre INSEE du bâti, le nombre de repérés, le nombre d'étudiés :

+, *ETUDIE*

exemple :

PART : grange ; colombier (étudié)

```
<PARTIES>
  <PART> grange </PART>
  <PART ETUDIE = «oui»> colombier </PART>
</PARTIES>
```

La règle est la décomposition du champ en éléments PART (séparateur ;), avec création d'un attribut ETUDIE en cas de présence de l'affixe *étudié*

l'affixe *étudié* devrait être complété par un lien vers le dossier correspondant.

champ SCLE

Ce champ contient les différentes datations des campagnes de construction (séparateur ;) : chaque campagne est caractérisée par un ou plusieurs descripteurs de datation (séparateur la virgule), avec la précision éventuelle pour chacune si elle a été détruite ou si elle est douteuse.

+, *DOUTE*, *DETRUIT*

exemple :

SCLE : 2e moitié du 12e siècle, 1e moitié du 13e siècle (?) ; 14e siècle (détruit)

```
<DATATION>
  <DATAT DOUTE = « ? »>
    <SCLE> 2e moitié du 12e siècle </SCLE>
    <SCLE> 1e moitié du 13e siècle </SCLE>
  </DATAT>
  <DATAT DETRUIT = «oui»>
    <SCLE> 14e siècle </SCLE>
  </DATAT>
</ DATATION >
```

La règle est la décomposition du champ en autant d'éléments DATAT que de séparateur ;. A l'intérieur de chaque segment : création d'un élément SCLE pour chaque date (séparateur ,). Les affixes (? et détruit) sont mis en attribut de l'élément DATAT.

champ ESCA

Ce champ est composé de segments (séparateur le ;) composés de trois parties : l'emplacement, puis (séparateur :) la forme et (séparateur ,) la structure de l'escalier :

+

exemple :

ESCA : escalier dans œuvre : escalier en fer à cheval, en charpente ; escalier hors œuvre : escalier en vis

```
<ESCALIERS>
  <ESCA>
    <EMPLA> escalier dans œuvre </EMPLA>
    <FORME> escalier en fer à cheval </FORME>
    <STRUCT> en charpente </STRUCT>
  </ESCA>
```

```
<ESCA>
  <EMPLA> escalier hors œuvre </EMPLA>
  <FORME> escalier en vis </FORME>
</ESCA>
</ESCALIERS>
```

La règle est la décomposition du champ entre autant d'éléments ESCA qu'il y a de segments séparés par des ; Chaque segment se décompose en 3 : première partie (jusqu'au :) dans l'élément EEMPLA, deuxième partie (jusqu'à la virgule) dans l'élément FORME, troisième partie dans l'élément STRUCT. A l'intérieur d'un même segment, toutes les parties ne sont pas forcément renseignées.

champ TECH

Ce champ contient les descripteurs correspondant aux décors portés, éventuellement précisés par l'affixe « étudié dans la base Palissy » :

+, ETUDPALISS

exemple :

TECH : menuiserie ; sculpture (étudiée dans la base Palissy)

```
<TECHNIQUES>
  <TECH> menuiserie </TECH >
  <TECH ETUDPALISS = «oui»> sculpture </TECH >
</TECHNIQUES>
```

La règle est la décomposition du champ en autant d'éléments TECH qu'il y a de descripteurs séparés par des ; L'affixe éventuelle (entre parenthèses) devient l'attribut de l'élément.

la base Illustration

La *direction des musées de France* possède principalement trois bases de données qu'elle souhaite reprendre dans l'outil de production de dossiers électroniques. Ces trois bases sont ARCHEOLOGIE, ETHNOLOGIE et JO-CONDE. Pour chacune de ces bases, le système descriptif se scinde en quatre groupes de champs, les champs monovaleur, les champs multivaleur, les champs multi-séparateurs et les champs à texte intégral.

- Les champs à texte intégral seront copiés tel quel dans un élément⁴ du schéma cible.
- Les champs monovaleur possède un seul terme qui alimente naturellement l'élément correspondant dans le schéma cible.
- Les champs multivaleur avec ou sans affixe doivent être découpés selon un séparateur prédéfini et rangés dans des éléments identiques sous un même terme générique dans le schéma cible.
- Les champs multi-séparateurs doivent être découpés en tenant compte d'une liste de séparateurs prédéfinis et rangés dans un fragment du schéma cible. Ces champs ont un traitement au cas par cas.

La base Joconde

La base Joconde contient 44 champs

- **Champs texte intégral**

13 sont des champs texte intégral sans traitement spécifique sauf une exception :

ADPT, APTN, ATTR, BIBL, COMM, DECV, DESC, DIMS, EXPO, PGEN, PINS, PREP, VENT*

- **champs monovalués**

4 sont des champs monovalués : DMAJ, DMIS, IMAGE, REF

- **champs multivaleurs sans affixe**

6 sont des champs multivaleurs sans affixe :

APPL, DOMN, GEOHI, REFIM, SREP, STYL

- **champs multivaleurs avec affixe**

10 sont des champs multivaleurs avec affixe (? ou numérique)

COPY, DACQ, DDPT, DENO, DREP, ECOL, LOCA, MILL, PERI, PEOC

- **champs multi-séparateurs**

11 sont des champs multi-séparateurs, à traiter au cas par cas

AUTR, DEPO, GENE, INSC, INV, PAUT, PAUT, PHOT, REPR, STAT, TITR, TECH

Explications champ par champ

- **13 sont actuellement des champs texte intégral sans traitement spécifique demandé ou possible :**

- ADPT vers <NOTE DEPOT> dans la DTD, volonté de l'intégrer dans l'élément <DEPO> avec l'attribut <ancien dépôt>
- APTN vers <NOTE COLLECTION> dans la DTD, volonté de structurer cet élément
- ATTR vers <NOTE AUTEUR> dans la DTD, volonté de l'intégrer dans l'élément <AUTR> avec l'attribut <ancienne attribution>
- BIBL vers <NOTE BIBLIOGRAPHIE> dans la DTD, volonté de structurer cet élément
- COMM vers <COMMENTAIRE>
- DECV vers <NOTE DECOUVERTE>
- DESC vers <DESCRIPTION>
- DIMS vers <NOTE DIMENSION> dans la DTD, volonté de structurer cet élément
- EXPO vers <NOTE EXPOSITION> dans la DTD, volonté de structurer cet élément
- PGEN vers <NOTE UTILISATION>
- PINS vers <NOTE INSCRIPTION>
- PREP vers <NOTE INDEXATION>

cas particulier

- VENT vers <NOTE COLLECTION> après le contenu du champ <APTN> s'il existe et précédé du terme «*Vente* : «

· **4 sont des champs monovalués :**

- DMAJ vers <NOTICE MODIFIEE>
- DMIS vers <NOTICE CREE>
- IMAGE vers <PRESENCE IMAGE>
- REF vers <REFERENCE>

<STYL>style2</STYL>
etc.
</STYLE MOUVEMENT>

• **10 sont des champs multivaleurs avec affixe ? ou numérique**

L'affixe ? doit être considéré comme un attribut de l'information (doute sur l'information)

➤ COPY

Ce champ est composé de parties (séparateur des virgules) : un ou plusieurs noms d'organisme et le dernier terme est toujours une date. La proposition serait la structure cible suivante :

+

exemple: COPY : (c) musée du Louvre, (c) direction des musées de France, 1999

```
<COPYRIGHT TEXTE>
<COPY>
  <COPY ORGANISME>(c) musée du Louvre </COPY ORGANISME>
  <COPY ORGANISME>(c) direction des musées de France </COPY ORGANISME>
  <COPY DATE>1999</COPY DATE>
</COPY>
```

➤ DACQ

La règle est la décomposition du champ en segments selon le séparateur point virgule. Suivant le contenu dans le segment le contenu doit aller vers <DATE ACQUISITION> ou <DATE ENTREE>

Chaque segment est ensuite décomposé en deux éléments en fonction du premier espace. Il peut y avoir qu'une seule valeur qui correspond à une date

Ce champ est composé d'une succession de segments (séparateur de segment le point-virgule). Chaque segment est composé d'une ou deux parties :

- si une partie, cela est une date qui va dans <DATE ACQU>

- si deux parties, le premier terme est une date (AAAA) correspondant à l'année connue, complétée par une précision de cette date (séparateur entre la date et type est l'espace). Si la précision contient le terme «entrée», le segment doit être placé dans <DATE ENTREE>

- la notion de doute existe et est traitée en attribut (ex : 1852 ; ?)

La proposition serait la structure cible suivante :

+

exemple : *DACQ* : 1898 date d'acquisition ; 1902 entrée matérielle

```
<DATE STATUT>
  <DATE ACQUISITION>
    <DACQ>
      <DATE ACQU>1898</DATE ACQU>
      <TYPE ACQU>date d'acquisition</TYPE ACQU>
    </DACQ>
  </DATE ACQUISITION>
  <DATE ENTREE>
    <DENT>
      <DATE ENTREE>1902</ DATE ENTREE>
      <TYPE ENTREE> entrée matérielle </TYPE ENTREE>
    </DENT>
  </DATE ENTREE>
</DATE STATUT>
```

➤ DDPT

Ce champ est composé d'une succession de segments (séparateur de segment le point-virgule). Chaque segment est composé d'une ou deux parties :

- si une partie, cela est une date (AAAA) qui va dans <DATE>
- si deux parties, le premier terme est une date (AAAA) correspondant à l'année connue, complétée par la précision de cette date (séparateur entre la date et type est l'espace).
- la notion de doute existe et est traitée en attribut (ex : 1852 ; ?)

La proposition serait la structure cible suivante :

+

exemple : *DDPT* : 1898 entre ; 1902 et

```
<DATE DEPOT>
  <DDPT>
    <DATE>1898</DATE>
    <AFFIXE>entre</AFFIXE>
  </DDPT>
  <DDPT>
    <DATE>1902</DATE>
    <AFFIXE> et </AFFIXE>
  </DDPT>
</DATE DEPOT>
```

➤ DENO

L'affixe ?, ou numérique ou précision doit être considéré comme un attribut de l'information (doute sur l'information, précision dénomination ou nombre d'objets). La proposition serait d'ajouter des attributs à l'élément décomposé.

Le champ DENO deviendrait un élément DENOMINATION décomposé en autant d'éléments DENO qu'il y a de termes, chaque élément DENO pouvant être affecté des attributs DOUTE, PRECISION ou NOMBRE

Exemple : *DENO* : assiette (ronde) ; paire (?)

```

<DENOMINATION>
  <DENO PRECISION = «ronde»>assiette</DENO >
  <DENO DOUTE = «?»>paire</DENO>
</DENOMINATION>

```

La règle de transformation doit permettre de mettre le contenu de la parenthèse en attribut et de choisir l'attribut cible selon la valeur de la parenthèse (? dans DOUTE, numérique dans NOMBRE et autre dans PRECISION).

➤ DREP

Ce champ est composé d'une succession de segments (séparateur de segment le point-virgule). Chaque segment est composé d'une ou deux parties :

- si une partie, cela est une date (AAAA ou AAA/MM/JJ) qui va dans <DATE REPR>
- si deux parties, le premier terme est une date (AAAA ou AAA/MM/JJ) correspondant à l'année connue, complétée par la précision de cette date (séparateur entre la date et le type est l'espace).

La proposition serait la structure cible suivante :

+

exemple : Buste de Cicéron => DREP : 106 AJC né ; 43 AJC mort

```

<DATE REPRESENTATION>
  <DREP>
    <DATE REPR>106 AJC</DATE REPR>
    <AFFIXE>né</AFFIXE>
  </DREP>
  <DREP>
    <DATE>43 AJC</DATE>
    <AFFIXE>mort </AFFIXE>
  </DREP>
</DATE REPRESENTATION>

```

La règle est la décomposition du champ en segments DREP selon le séparateur point virgule. Chaque segment est ensuite décomposé en deux éléments en fonction du premier espace SAUF si la date est suivie de AJC . Il peut y avoir qu'un seul terme qui correspond forcément à une date.

➤ ECOL

La règle est la décomposition du champ en segments selon le séparateur point virgule. Chaque segment est ensuite décomposé en plusieurs éléments séparés par des virgules.

Comme pour DENO, on peut utiliser l'affixe ? entre parenthèses, qui doit être considéré comme un attribut de l'information (doute sur l'information)

Exemple : France, Rouen (?) ; Grande-Bretagne (?)

```

➤ ECOL deviendrait
  <ECOLE>
    <ECOL>France</ECOL>
    <ECOL DOUTE = «?»>Rouen</ECOL>
  </ECOLE>
  <ECOLE>
    <ECOL DOUTE = «?»>Grande-Bretagne</ECOL>

```

etc.
</*ECOLE*>

➤ **LOCA**

Ce champ contient pour la majorité des notices deux termes séparés par un point virgule.

Cas général :

Le 1er est toujours le nom d'une ville et doit aller dans <LOCA VILLE>

Le 2e terme est toujours le nom d'un établissement et doit aller dans <LOCA ETABLISSEMENT>

Cas particulier

Quelques notices correspondent à des oeuvres disparues avec une précision : «localisation inconnue», «œuvre disparue» ou «œuvre détruite».

- Le terme <Localisation inconnue> est le seul terme. Il remplace le cas général.

- Les termes «œuvre disparue» ou «œuvre détruite» doivent être traités comme des attributs.

<LOCALISATION>
 <LOCA VILLE>ville</LOCA VILLE>
 <LOCA ETABLISSEMENT>oeuvre disparue>établissement</LOCA ETABLISSEMENT>
</LOCALISATION>

ou

<LOCALISATION>
 <LOCA INCONNUE>localisation inconnue</LOCA INCONNUE>
</LOCALISATION>

➤ **MILL**

Même traitement que pour le champ DDPT

Ce champ est composé d'une succession de segments (séparateur de segment le point-virgule). Chaque segment est composé d'une ou deux parties :

- si une partie, cela est une date qui va dans <DATE>

- si deux parties, le premier terme est une date (AAAA) correspondant à l'année connue, complétée par la précision de cette date (séparateur entre la date et type est l'espace).

La proposition serait la structure cible suivante :

+

exemple : *MILL* : 1898 entre ; 1902 et

<MILLESIME>
 <MILL>
 <DATE>1898</DATE>
 <AFFIXE>entre</AFFIXE>
 </MILL>
</MILL>

```
<DATE>1902</DATE >
<AFFIXE> et </AFFIXE >
</MILL>
</MILLESIME>
```

➤ PEOC et PERI

La règle est la décomposition de ces deux champs en segments selon le séparateur point virgule. Chaque segment contient un élément qui peut être affecté de l'attribut DOUTE. L'affixe ? entre parenthèses, doit être considéré comme un attribut de l'information (doute sur l'information).

Exemple : PEOC : 3e quart 18e siècle (?)

```
<ORIGINAL COPIE>
  <PEOC DOUTE = «?»>3e quart 18e siècle</PEOC>
</ORIGINAL COPIE>
```

Exemple : PERI : 3e quart 18e siècle (?)

```
<SIECLE>
  <PERI DOUTE = «?»>3e quart 18e siècle</PERI>
</SIECLE>
```

• **champs multi-séparateurs**

11 sont des champs multi-séparateurs, à traiter au cas par cas
AUTR, DEPO, GENE, INSC, INV, PAUT, PHOT, REPR, STAT, TITR, TECH

➤ AUTR

La règle est la décomposition du champ en segments AUTR selon le séparateur point virgule. Chaque segment est ensuite décomposé en plusieurs éléments en fonction de différents séparateurs : la virgule, les parenthèses, l'espace.

Actuellement, ce champ regroupe un ensemble d'informations différentes que l'on peut identifier :

Pour chaque segment, la règle générale est la suivante :

- Les descripteurs en majuscules sont mis dans l'élément NOM, le(s) suivant(s) dans l'élément PRENOM, Les descripteurs entre parenthèses sont de différentes natures.

Cas particuliers à traiter :

- présence d'un complément au nom : " fils ", rang dans la hiérarchie... (présent dans le segment entre le prénom et la parenthèse de la profession) : créer un élément COMPL,

- présence d'un pseudonyme (présent dans le segment entre le prénom et la parenthèse de la profession) : terme introducteur " , dit " ou " ,dite " : créer un élément PSEUDO, décomposé lui aussi en NOM et PRENOM,

- présence d'un nom de jeune-fille : (présent dans le segment entre le prénom et la parenthèse de la profession) : terme introducteur " , née ", créer un élément NOMJF,

1/ **le nom de l'auteur est en majuscules suivi des prénoms en minuscules** séparés par des espaces.

Exemple : RUBENS Pierre Paul

```
<AUTR>
  <NOM> RUBENS </NOM>
  <PRENOM> Pierre Paul </PRENOM>
</AUTR>
```

2/ **Cas des oeuvres anonymes et utilisation d'attributs**

La présence entre les parenthèses des termes " école de ", " atelier de ", «d'après» est à mettre dans l'attribut TYPE.

Exemple : Anonyme ; LE BRUN Charles (d'après)

```
<AUTR>
  <NOM> Anonyme </NOM>
</AUTR>
<AUTR>
  <NOM> LE BRUN </NOM>
  <PRENOM> Charles </PRENOM>
  <AUTRTYPE = «d'après»>
</AUTR>
```

3/ éventuellement des précisions sont mentionnées entre parenthèses, il peut y en avoir plusieurs séparées par des virgules :

Exemple : RUBENS Pierre Paul (? , école)

```
<AUTR DOUTE= « ? »>
  <NOM> RUBENS </NOM>
  <PRENOM> Pierre Paul </PRENOM>
  <AUTRTYPE = « école »>
</AUTR>
```

4/ Dans le même segment, on peut avoir la mention du **pseudonyme** séparé par une virgule après le nom et prénom. La présence d'une particule dans le nom est à rejeter dans l'élément PARTIC

Exemple : RIJN REMBRANDT Harmensz. Van, REMBRANDT (dit, école)

```
<AUTR>
  <NOM>RIJN REMBRANDT</NOM>
  <PRENOM>Harmensz.</PRENOM>
  <PARTIC>Van</PARTIC>
  <PSEUDO>
    <NOM>REMBRANDT</NOM>
  </PSEUDO>
</AUTR>
```

Quelques problèmes rencontrés : **le traitement de la particule** (actuellement rejetée après le prénom, comme il se doit dans les normes de citation) : si on scinde nom et prénom, on ne peut la rattacher au prénom : ou bien on la rapatrie après le nom (mais impossibilité de remise en forme automatique après le prénom pour les publications...) ou bien on crée un élément indépendant PART, **le traitement des numéros de dynastie ou des “ le jeune ”, “ le vieux ”...**

5/ Le rôle de l'artiste n'est pas systématiquement indiqué en particulier pour les peintures et les sculptures. Par contre on précise le rôle entre parenthèses pour des intervenants sur l'œuvre de l'artiste :

Exemple : BARDEY Jeanne ; RUDIER Alexis (fondeur)

```
<AUTEUR>
  <AUTR>
    <NOM>BARDEY </NOM>
    <PRENOM>Jeanne</PRENOM>
  </AUTR>
  <AUTR>
    <NOM>RUDIER </NOM>
    <PRENOM>Alexis</PRENOM>
    <PROFESSION>fondeur</PROFESSION >
  </AUTR>
</AUTEUR>
```

+

+

exemple

AUTR : COUSIN Jean le Fils (?) ; GELLE Claude, LE LORRAIN (dit) ; FANTIN-LATOURE Victoria, DUBOURG (née) ; BALDUNG Hans, BALDUNG Grien (dit, ? , école)

```

<AUTEUR>
  <AUTR DOUTE= « ? »>
    <NOM> COUSIN </NOM>
    <PRENOM> Jean </PRENOM>
    <COMPL> le Fils </COMPL>
  </AUTR>
  <AUTR>
    <NOM> GELLEE </NOM>
    <PRENOM> Claude </PRENOM>
    <PSEUDO>
      <NOM> LE LORRAIN </NOM>
    </PSEUDO>
  </AUTR>
  <AUTR>
    <NOM> FANTIN-LATOURE </NOM>
    <PRENOM> Victoria </PRENOM>
    <NOMJF> DUBOURG </NOMJF>
  </AUTR>
  <AUTR DOUTE= « ? » TYPE = « école »>
    <NOM> BALDUNG </NOM>
    <PRENOM> Hans </PRENOM>
    <PSEUDO>
      <NOM> BALDUNG </NOM>
      <PRENOM> Grien </PRENOM>
    </PSEUDO>
  </AUTR>
</AUTEUR>

```

➤ DEPO deviendrait

Ce champ contient trois termes séparés par un point virgule.

Le 1er terme est toujours le type de dépôt et doit aller dans <TYPE DEPOT>

Le 2e terme est toujours le nom d'une ville et doit aller dans <DEPOT VILLE>

Le 3e terme est toujours le nom d'un établissement et doit aller dans <DEPOT ETABLISSEMENT>

```

<DEPOT>
  <TYPE DEPOT> type dépôt </TYPE DEPOT>
  <DEPOT VILLE> ville dépôt </DEPOT VILLE>
  <DEPOT ETABLISSEMENT> établissement dépôt </DEPOT ETABLISSEMENT>
</DEPOT>

```

champ GENE

Ce champ est composé d'une ou plusieurs séquences (séparateur point-virgules). Une séquence peut contenir un ou plusieurs termes recouvrant des informations différentes. Pour répartir correctement les informations, il faut identifier le contenu de ce champ dans un cas particulier.

Le contenu se fait dans le cadre de "phrases" (séparateur de phrase : le point-virgule ;) correspondant à une signification. La phrase est composée de mots-clés, agencés selon une syntaxe faisant intervenir les séparateurs parenthèses (), virgule et deux-points.

Cas particulier

1/ Type d'utilisation : présence du terme «**utilisation entre parenthèses**» le contenu doit aller dans <TYPE UTILISA-

TION> sans la reprise du terme «(utilisation)»
ex GENE : service-consommation des aliments (utilisation)

```
<TYPE UTILISATION>  
<TYPE UTIL>service-consommation des aliments</TYPE UTIL>  
</TYPE UTILISATION>
```

2/ Dans la majorité des cas

La règle est la décomposition du champ en autant de segments séparés par des ;

Pour chacun de ces segments, on crée autant d'éléments GENESE qu'il y a de " phrases " séparées par des ;

Enfin, chaque descripteur (séparateur parenthèses, virgules et deux-points) génère un élément GENE.

+

+

Question : faut-il encore aller plus loin dans la modélisation de la " phrase " : faut-il la voir comme ci-dessus comme une succession de mots-clés (descripteurs pris dans une liste fermée à indexer) ou comme une simple phrase (texte libre à indexer en texte intégral) ou aller encore plus loin dans l'explicitation de la sémantique des séparateurs () , et : ?

exemple

GENE : réplique ; Angleterre (édifice commercial, provenance) ; oeuvre en rapport

Précision : peinture ayant servi d'enseigne à un marchand Anglais

```
<GENESE >  
  <GENE>  
    <MOTCLE>réplique</MOTCLE>  
  </GENE>  
  <GENE>  
    <MOTCLE> Angleterre </MOTCLE>  
    <MOTCLE> édifice commercial</MOTCLE>  
    <MOTCLE> provenance</MOTCLE>  
  </GENE>  
  <GENE>  
    <MOTCLE> oeuvre en rapport </MOTCLE>  
  </GENE>  
</GENESE>
```

Il faut encore définir le traitement spécifique des champs INSC, INV, PAUT, PHOT, REPR, STAT, TITR, TECH. Les traitements sont relativement similaires aux précédents.

1.6 Journalisation des erreurs

Les documents n'ayant pas été importés sont indiqués dans un fichier journal. Pour chaque document en erreur, le journal indique l'identification du document et la cause de l'erreur (manque d'informations pour créer le document, type incompatible, dépassement de capacité, liens non résolus).

¹ Si une solution de stockage sans paramétrage ne peut être appliquée,

² voir la DTD biblioML pour plus de précision.

³ le nom de l'élément cible n'est pas forcément le même que celui du champ d'origine. Il se peut que l'objet cible soit un attribut et non un élément.

⁴ le nom de l'élément cible n'est pas forcément le même que celui du champ d'origine. Il se peut que l'objet cible soit un attribut et non un élément.

