



Numérisation et encodage en EAD d'instruments de recherche

Recommandations et aide à la rédaction
du cahier des charges

Janvier 2015

Note sur le présent document

Rédigé par les Archives de France (sous-direction de la politique archivistique, bureau de l'archivage numérique, des normes et référentiels ; sous-direction de la communication et de la valorisation des archives, bureau de l'accès aux archives et de la diffusion numérique), ce document a pour but de proposer aux services souhaitant s'engager dans la rétroconversion au format XML/EAD d'instruments de recherche, une assistance à la rédaction d'un cahier des charges pour l'encodage. Il s'agit d'un cahier des charges type. Seuls les aspects métier sont abordés ici.

Ce modèle de cahier des charges est diffusé en complément de l'appel à projets lancé par les Archives de France pour l'encodage des instruments de recherche.

Les recommandations qui figurent dans ce document sont celles qui apparaissent aux Archives de France comme garantissant le meilleur encodage possible compte tenu de la diversité des instruments de recherche à rétroconvertir. Des spécifications sont données pour une mise en conformité avec le profil européen de l'EAD (apeEAD) et par rapport aux évolutions prévues de l'EAD elle-même (révision en cours par la Société des archivistes américains). Chaque service disposant de ses propres outils de publication, la question de l'intégration, dans les différents systèmes d'information des archives départementales, communales ou intercommunales, des instruments de recherche ainsi rétroconvertis ne sera pas abordée ici.

Les Archives de France tiennent à remercier tout particulièrement les archives départementales des Yvelines (M^{me} Catherine Junges, directeur adjoint), du Doubs (M^{me} Rachel Froissart, responsable du pôle Collecte), du Pas-de-Calais (M. Lionel Gallois, directeur) et de la Marne (M^{me} Manonmani Restif, adjointe au directeur), pour l'envoi d'exemples de cahiers des charges.

Le bureau de l'archivage numérique, des normes et référentiels (Claire Sibille - de Grimouard - claire.sibille@culture.gouv.fr et Alice Motte – alice.motte@culture.gouv.fr) et le bureau de l'accès aux archives et de la diffusion numérique (Lucile Grand – lucile.grand@culture.gouv.fr) se tiennent à la disposition des services d'archives, notamment pour toute question sur le contenu du document.

Lecture du cahier des charges type

Lorsque des remarques à l'attention du service sont faites, le texte est en décalé par rapport au corps de texte principal, en italique et en bleu.

Exemple de texte à destination du service d'archives, pour expliciter certains points.

Cahier des charges type pour la numérisation et l'encodage en EAD d'instruments de recherche

Table des matières

1. Définition du projet et de la prestation.....	4
1.1. Contexte du projet.....	4
1.2. Périmètre de la prestation.....	5
2. Présentation des instruments de recherche à numériser et à encoder.....	6
2.1. Description des instruments de recherche.....	6
2.2. Préparation et récolement par le service d'archives.....	6
3. Modalités de fonctionnement.....	7
3.1. Conduite du projet.....	7
3.1.1. Réunions.....	7
3.1.2. Démarche qualité.....	8
3.2. Calendrier de la prestation et échéancier.....	8
4. Lot 1 – Numérisation d'instruments de recherche.....	8
4.1. Description et déroulement de la prestation.....	8
4.2. Description des différentes phases.....	9
4.3. Règles pour le nommage des fichiers de données textuelles.....	10
4.4. Description des modalités de contrôle qualité de l'océrisation.....	11
5. Lot 2 – Structuration et encodage d'instruments de recherche.....	11
5.1. Description des instruments de recherche faisant l'objet de la prestation.....	11
5.2. Description et déroulement de la prestation.....	12
5.3. Description des différentes phases.....	13
5.4. Règles pour le nommage des fichiers produits.....	22
5.5. Description des modalités de contrôle qualité de l'encodage.....	22
Annexe 1 - Exemple de grille descriptive d'instruments de recherche.....	24
Annexe 2 - Attributs à ajouter à certains éléments.....	25
Annexe 3 – Valeurs de l'attribut @normal de l'élément <unitdate>.....	28

Cahier des charges type pour la numérisation et l'encodage en EAD d'instruments de recherche

1. Définition du projet et de la prestation

1.1. Contexte du projet

Le Service interministériel des Archives de France, avec la volonté d'impulser et d'accélérer la dématérialisation des ressources archivistiques ainsi que leur visibilité sur le Web, a lancé un nouvel appel à projet pour la numérisation et l'encodage en XML/EAD des instruments de recherche disponibles dans les archives départementales et communales sur plusieurs types de fonds :

- Les fonds en relation avec la guerre de 1914-1918. Pour les archives départementales, les instruments de recherche de la série R (affaires militaires), ainsi que les bordereaux de versement d'archives relatives aux affaires militaires (« suite » de la série R). Pour les archives (inter)communales, les instruments de recherche des archives relatives aux affaires militaires, sans distinction chronologique ;
- Les fonds privés et les fonds iconographiques dont les instruments de recherche, par leur richesse, leur diversité et leur description très poussée, sont particulièrement adaptés à la diffusion et à la recherche en ligne ;
- Les fonds les plus sollicités et accessibles au public le plus large. Pour les archives départementales, les instruments de recherche des séries M (administration générale), O (administration et comptabilité communale), Q (domaines, enregistrement et hypothèques), E-dépôt (archives communales), E (notaires), U (justice), ainsi que les bordereaux de versements d'archives de l'administration générale (« suite » de la série M). Pour les archives (inter)communales, les instruments de recherche des archives de l'administration générale, sans distinction chronologique.

Les instruments de recherche ainsi encodés ont vocation à être publiés sur les sites des archives départementales ou (inter)communales puis, une fois publiés, à être moissonnés pour intégration dans des portails documentaires nationaux ou internationaux, archivistiques ou patrimoniaux.

Ce projet rejoint également certaines des recommandations du rapport de M. Maurice Quénet du 8 avril 2011 sur l'avenir des Archives de France, suggérant de regrouper virtuellement les fonds conservés dans les services d'archives français au sein d'un portail interministériel unifié d'accès aux instruments de recherche et aux fonds d'archives numérisées, dont le lancement est prévu à l'horizon 2016, ayant lui-même vocation à alimenter le portail archivistique européen.

Il est souhaité également à terme de relier ces ressources archivistiques à d'autres ensembles de données, dans une perspective de Web sémantique dont le bénéfice sera d'améliorer la visibilité des archives. L'encodage constitue un premier palier dans cet enjeu important.

1.2. Périmètre de la prestation

Le cahier des charges envisage deux lots selon que les instruments de recherche à encoder sont manuscrits, dactylographiés, imprimés ou déjà au format numérique :

a) lot 1 : numérisation d'instruments de recherche par reproduction en photographie numérique, océrisation en mode texte et reprise éventuelle de la saisie ;

b) lot 2 : structuration et encodage d'instruments de recherche.

Le service conservera les clauses adéquates en fonction de son contexte et de ses choix en la matière. Les instruments de recherche peuvent aussi être composites et comporter des parties et éléments sur des supports et dans des formats divers (pages dactylographiées, notes manuscrites, imprimés, fichiers bureautiques et bases de données). Il convient dans ce cas de le préciser clairement lorsque l'on décrira les documents soumis à la numérisation.

La prestation attendue consiste à :

Lot 1 :

- numériser des instruments de recherche manuscrits, dactylographiés et imprimés des fonds précités, appliquer un traitement d'océrisation en mode texte, contrôler l'océrisation et reprendre éventuellement la saisie ;
- produire des fichiers bureautiques DOC ou ODT ;
- livrer au service d'archives les fichiers numériques DOC ou ODT ainsi produits selon des modalités qui seront définies ci-après (livraison par supports amovibles ou par un protocole de transfert réseau comme FTP).

Lot 2 :

- rendre des instruments de recherche conformes à la *Norme générale internationale de description archivistique* ou norme ISAD(G) et les convertir au format XML conformément à la DTD EAD 2002 (Description archivistique encodée), dont la version officielle est diffusée sur le site de la Bibliothèque du Congrès <http://lcweb.loc.gov/ead/> ;
- fournir des feuilles de transformation XSL permettant la génération de fichiers PDF et/ou HTML à partir des fichiers XML/EAD ;
- convertir les fichiers EAD 2002 en fichiers conformes au profil apeEAD élaboré pour le Portail européen des archives (<http://www.archivesportaleurope.net/>) au moyen de l'outil de conversion fourni dans le cadre du projet (<http://www.apenet.eu/> Rubrique " A propos d'APEnet / Développements ").
- livrer au service d'archives les fichiers numériques ainsi produits (fichiers XML EAD 2002 et apeEAD, feuilles de style XSL) selon les modalités de livraison définies ci-après (livraison par supports amovibles ou par un protocole de transfert réseau comme FTP).

2. Présentation des instruments de recherche à numériser et à encoder

2.1. Description des instruments de recherche

Les instruments de recherche originaux, au nombre de XXX, sont disponibles sur support papier, pour un total de XXX pages. Ils se présentent sous forme imprimée, dactylographiée, voire manuscrite, et ils peuvent être brochés ou reliés. Le format de chaque page est en général légèrement supérieur à l'A4. Selon les époques, la qualité du papier est variable (papier épais, papier jauni, papier cassant, papier froissé, etc.). Les périodes, les qualités diverses d'impression et la variété des typographies peuvent éventuellement causer des problèmes en cas d'application d'un traitement d'océrisation (lettres non ou mal reconnues, lacunes, coquilles, césures de mots). Une attention particulière devra y être portée. Le prestataire pourra également avoir à tenir compte de la présence de retombes comportant des informations complémentaires ou d'annotations manuscrites, sur les instruments de recherche tapuscrits ou imprimés.

Un instrument de recherche complet est composé théoriquement de plusieurs parties définies par la réglementation archivistique : une introduction historique, une bibliographie, des sources complémentaires, un plan de classement, le corps de l'instrument de recherche structuré de façon hiérarchique, un ou plusieurs index, un tableau de concordance des cotes, une ou plusieurs annexes et une table des matières.

La forme des instruments de recherche varie dans le temps et selon le type de classement effectué. Certains ne comportent que le corps de l'instrument de recherche, d'autres comportent des notes de bas de page. Leur structure est donc très hétérogène et demande un travail très fin de lecture et d'analyse.

Le service ajustera au besoin cette description des instruments de recherche à numériser selon les informations qu'il a relevées lors du récolement et de la préparation des documents. Les informations doivent idéalement être consignées dans le fichier de récolement dressé lors de la préparation des documents (voir paragraphe 2.2) et dans la grille descriptive détaillée (voir annexe 1).

2.2. Préparation et récolement par le service d'archives

Le service d'archives mettra à disposition du prestataire une série d'instruments de recherche. Selon la nature des supports et l'état physique des instruments de recherche, cet ensemble pourra être découpé en lots homogènes d'instruments de recherche. Les cas problématiques devront être repérés pour faciliter l'analyse des résultats de l'OCR. Chaque lot livré pour la numérisation pourra faire l'objet d'un bon de commande.

Les documents livrés auront été récolés et préparés par le service d'archives en vue de leur numérisation. Le service d'archives en contrôlera la cohérence d'ensemble ainsi que les pages manquantes. Les opérations de massicotage et de démontage de reliures auront été préalablement effectuées par le service qui fera également reproduire les instruments de recherche pour transmettre au prestataire les originaux et en garder des copies.

Pour chaque lot de documents à numériser, il sera fourni au prestataire un fichier de récolement issu du travail d'analyse préalable. Ce fichier contiendra les informations permettant au prestataire d'identifier les documents lors de leur enlèvement et de leur réception dans les locaux de numérisation. Il indiquera les particularités et problèmes potentiels posés par chaque instrument de recherche.

Le fichier de récolement devra également déterminer la granularité attendue de la numérisation et de l'encodage (un instrument de recherche papier correspondant à un seul instrument de recherche au format numérique, ou plusieurs instruments de recherche au format papier fusionnés pour donner un seul instrument de recherche numérique ou encore un instrument de recherche papier pour plusieurs instruments de recherche numériques).

A priori, le cas le plus simple consistera à prendre la sous-série du cadre de classement comme niveau de granularité.

Le fichier de récolement pourra être constitué des colonnes suivantes :

Série	Sous-série	Auteur Titre Tome Année d'édition	Format en cm ouvert	Support matériel	Nombre de pages total	Problèmes potentiels (reliure, impression, complexité de mise en page ou de repérage des éléments...)	Cible numérique ¹

3. Modalités de fonctionnement

3.1. Conduite du projet

Le prestataire retenu devra désigner un chef de projet qui sera l'interlocuteur direct du chef de projet du service d'archives.

3.1.1. Réunions

Le projet commencera par la tenue d'une réunion de cadrage entre le prestataire et le service d'archives, représentés chacun par le chef de projet, afin de convenir à la fois des aspects organisationnels, administratifs et techniques.

Des réunions de suivi devront être tenues à échéances régulières, par exemple à l'issue du traitement d'un bon de commande. Il conviendra de définir la logique calendaire de ces réunions lors de la réunion de cadrage.

Selon les étapes de la commande, les réunions seront destinées à examiner conjointement les instruments de recherche pour en préparer les spécifications détaillées, faire un point d'étape sur les travaux en cours chez le prestataire, présenter les livrables et préparer les demandes de correction après contrôle des livrables par le service d'archives.

Chaque réunion devra faire l'objet d'un compte rendu rédigé par le chef de projet du

¹ Préciser si les instruments de recherche sont à maintenir tels quels, à fusionner ou à éclater.

prestataire.

3.1.2. Démarche qualité

Au début du projet, le prestataire établira un plan de qualité de projet qui sera validé par le service d'archives. Ce document de référence définira les modalités de mise en œuvre des opérations de numérisation et d'encodage, le mode de fonctionnement entre le prestataire et le service d'archives, et devra résumer à tout le moins :

- les moyens humains, matériels et techniques qui seront mobilisés par le prestataire pour réaliser la prestation ;
- les modalités d'organisation entre le service d'archives et le prestataire, précisant notamment, dans le déroulement général de la prestation, l'enchaînement de la livraison des documents ou des lots de documents, la production des données, le contrôle qualité et la validation des lots, avec les délais associés à ces phases et le rythme de livraison des lots de documents ;
- les paramètres techniques retenus pour la numérisation et l'océrisation des instruments de recherche ;
- les règles de mise en conformité avec la norme ISAD(G) et les règles d'encodage des instruments de recherche ;
- les supports de livraison des données pour le contrôle qualité et la livraison définitive des données validées ;
- les règles de nommage des fichiers et, le cas échéant, des supports amovibles de livraison ;
- les procédures de contrôle qualité appliquées par le prestataire et les critères définis pour le contrôle qualité dont conviendront le service d'archives et le prestataire ;
- la forme et le contenu des rapports de production livrés par le prestataire à chaque lot.

Le prestataire devra soumettre au service d'archives, au fur et à mesure qu'ils se présenteront, les anomalies ou problèmes non résolus : ceux-ci doivent d'une part être notifiés au service dès leur apparition, par voie téléphonique ou par messagerie, afin de trouver une solution au plus vite, d'autre part être consignés dans le rapport de production qui sera livré par le prestataire si le problème s'avère insoluble.

3.2. Calendrier de la prestation et échéancier

Le calendrier et l'enchaînement des différentes étapes de l'opération (récupération des instruments de recherche, production des données, contrôle qualité, livraison des données, retour des instruments de recherche) seront définis et ajustés lors de la réunion de cadrage avec le prestataire du marché.

Nota : soit le service livre au prestataire les instruments de recherche à numériser et à encoder, soit le prestataire vient récupérer le lot aux archives. Il conviendra dans tous les cas de vérifier que le prestataire a bien tout reçu sur la base du fichier de récolement.

4. Lot 1 – Numérisation d'instruments de recherche

4.1. Description et déroulement de la prestation

La prestation attendue est la suivante :

- selon les modalités définies entre le service d'archives et le prestataire du marché, récupération par le prestataire des instruments de recherche papier et vérification de la complétude de la livraison sur la base du fichier de récolement détaillé fourni par le service d'archives avec le lot ;
- numérisation des instruments de recherche papier en mode image ;
- traitement par OCR (reconnaissance optique de caractères) pour générer un document en mode texte ;
- contrôle de la qualité de l'OCR par le prestataire et reprise éventuelle de la saisie des instruments de recherche indiqués ci-dessus ;
- formatage du document texte en fichiers bureautiques de type DOC ou ODT ne tenant pas compte de la mise en page.

Autre scénario possible : des instruments de recherche peuvent avoir été numérisés antérieurement. Des tests devront être réalisés avant le lancement de l'opération pour évaluer leur qualité et déterminer si le prestataire doit repartir des instruments de recherche papier initiaux pour les opérations de numérisation.

4.2. Description des différentes phases

Phase 1 : les instruments de recherche papier fournis par le service d'archives seront convertis en fichiers numériques reprenant l'intégralité des informations contenues dans la version papier.

Le service d'archives souhaite s'appuyer sur les critères techniques suivants. Toutefois, ces critères pourront être éventuellement adaptés lors de la phase de test avec le prestataire. En cas de réévaluation des critères techniques, un nouveau document consignait ces critères sera produit et servira de base pour la prestation et le contrôle qualité effectué par le service d'archives.

Numérisation en mode image

Le prestataire commencera par produire une image du document qui servira de base pour le traitement de l'OCR. Il est souhaité une image présentant les caractéristiques suivantes :

- Rapport d'agrandissement : 1/1
- Format : TIFF v6 (de préférence non compressé)
- Échantillonnage : 8 bits (256 niveaux de gris)
- Résolution de restitution : 400 dpi
- Post-traitement éventuel : redressement de l'image, détournage autour du bloc de texte, effacement des pinces ou autres marques physiques, renforcement du contraste et de la netteté (à utiliser toutefois à précaution pour ne pas générer d'artefacts gênants).

Traitement par OCR

Sur la base des images générées, le prestataire appliquera un traitement OCR exhaustif. Le format texte obtenu doit être encodé en UTF-8. Il lui est demandé à l'issue du traitement de vérifier la qualité de l'OCR.

Tout ou partie de certains instruments de recherche peut se présenter sous forme manuscrite, tapuscrite ou imprimée de telle sorte qu'un OCR peut s'avérer impossible (caractères difficilement

reconnaissables ou trop petits). Des éléments manuscrits complémentaires ou rectificatifs ont pu également être apportés au fil des ans sur les instruments de recherche imprimés.

Ce repérage des cas problématiques sera précisé par le service d'archives dans un fichier de récolement qui accompagnera chaque lot d'instruments de recherche confié au prestataire. Dans l'hypothèse où un cas problématique n'aurait pas été décelé par le service, le prestataire le signalera immédiatement au commanditaire par voie téléphonique ou par courriel, afin de convenir de la suite à donner au traitement.

Face à ces cas problématiques, une prestation de saisie des données, préalable à l'encodage, sera alors effectuée par le prestataire. Le choix de recourir à la saisie plutôt qu'à la numérisation et à l'océrisation sera déterminé sur proposition du prestataire et en accord avec le service.

Nota : la durée du contrôle qualité par le service d'archives ne doit pas être sous-estimée.

Phase 2 : elle débutera après la validation par le service d'archives des fichiers issus de l'OCR.

Les fichiers au format DOC ou ODT livrés avec leur feuille de style respecteront la structuration archivistique des instruments de recherche (c'est-à-dire l'enchaînement des différents composants de l'instrument) sans tenir compte de la mise en page (par exemple double colonne).

Phase 3 : livraison aux archives des fichiers numériques DOC ou ODT et du rapport de production.

Les fichiers numériques aux formats DOC ou ODT seront livrés au service d'archives sur le support qu'il aura souhaité.

Le prestataire livrera également un rapport de production qui complétera le fichier de récolement initialement fourni par les archives et comportant les informations suivantes :

- identifiants des fichiers texte ;
- rapport de production (date de production des données, incidents de production, défauts éventuels et toute information utile concernant les difficultés de réalisation, taux de qualité de l'OCR constaté et garanti par le prestataire). Il est à noter que ce document pourra servir aussi lors de la phase de contrôle qualité par les archives : des colonnes indiquant la date et le résultat du contrôle qualité par le commanditaire pourront ainsi enrichir l'historique de la production.

4.3. Règles pour le nommage des fichiers de données textuelles

Les fichiers seront nommés conformément au modèle suivant, préconisé par le ministère de la Culture et de la Communication (Département de la recherche, de l'enseignement supérieur et de la technologie) :

- FRAD000 : code normalisé de l'institution désignant le service d'archives ;
- code opération (sur quatre chiffres) interne au service d'archives ;
- identifiant de la série correspondant à l'instrument de recherche encodé ;

- numéro d'ordre de chaque fichier, si besoin, suite numérique sur deux caractères uniquement ;
- EXT : extension déterminant le format physique du fichier (ici DOC ou ODT).

Exemple : FRAD018_0008_1M_01.doc

4.4. Description des modalités de contrôle qualité de l'océrisation

Le contrôle de la qualité consistera à vérifier :

- l'intégralité de la prestation (complétude de la livraison avec la présence des fichiers DOC ou ODT par rapport au nombre d'instruments de recherche initiaux) ;
- le remplissage du fichier de récolement pour le rapport de production et la présence de tous les éléments demandés (identifiants des fichiers) ;
- la lisibilité des fichiers et leur exploitabilité ;
- l'encodage des caractères ;
- la cohérence (respect des règles de nommage énoncées au paragraphe 4.3) ;
- le contrôle intégral du taux d'exactitude de l'OCR sur les instruments de recherche ; le contrôle du respect de la mise en page initiale. Pour l'OCR, est considéré comme erreur tout mot erroné quel que soit le nombre de caractères erronés qu'il contient. Ne seront pas comptabilisées les erreurs de ponctuation sauf si un caractère de ponctuation est reconnu à la place d'une lettre ou bien encore les erreurs d'accent ;
- l'exactitude de la structuration des fichiers DOC ou ODT en respect des préconisations mises au point au début de la phase 2.

Le contrôle du respect des spécifications élaborées pendant la phase 2 pour les fichiers DOC ou ODT portera sur un échantillon de 20 % des pages du corps des instruments de recherche. Il consistera à vérifier que toutes les préconisations mises au point ont été respectées.

Le taux d'erreurs admissibles est fixé à 1,5 % maximum par document par rapport au nombre total de mots. Un instrument de recherche peut avoir des parties dont le taux de reconnaissance est supérieur ou inférieur au taux d'erreurs admissibles, mais la moyenne de l'ensemble doit correspondre à la qualité exigée.

5. Lot 2 – Structuration et encodage d'instruments de recherche

5.1. Description des instruments de recherche faisant l'objet de la prestation

La prestation porte sur l'encodage des instruments de recherche suivants :

- ...
- ...
- ...

Différents scénarios sont possibles.

Scénario 1 :

Le service d'archives fournit au prestataire les fichiers DOC ou ODT issus des opérations de numérisation et d'océrisation du lot 1.

Scénario 2 :

Les instruments de recherche existent déjà sous forme électronique, au format traitement de texte, tableur, base de données (contenu d'un progiciel documentaire, mais aussi de type ACCESS ou MySQL), ainsi qu'au format XML. Les données à structurer et à normaliser seront remises par le service d'archives au prestataire sous forme électronique (fichiers XML, fichiers EXCEL ou ODS, fichiers DOC ou ODT, avec ou sans feuille de style).

Scénario 3 :

Il peut être demandé au prestataire de réaliser une mise à jour d'instruments de recherche déjà encodés en XML conformément à l'EAD 2002, qu'ils aient été encodés dans le cadre du présent marché ou dans un autre contexte. Les mises à jour peuvent notamment consister en :

- *l'intégration de la description de nouvelles unités documentaires (ajouts d'éléments <c> et de leurs composants) ou la complétion de la description d'unités documentaires existantes (ajouts d'éléments dans des éléments <c>);*
- *la modification de balises existantes (modification de leur contenu, suppression de certaines balises et de leur contenu, transfert du contenu d'une balise vers une autre, suppression, modification ou ajout d'attributs et de leurs valeurs, etc.).*

5.2. Description et déroulement de la prestation

Lot 2 : structuration conformément à la norme ISAD(G) et transformation au format XML, suivant la DTD EAD 2002 (Description archivistique encodée), dont la version officielle est diffusée sur le site de la Bibliothèque du Congrès (<http://lcweb.loc.gov/ead/>).

- Phase 1 : structuration conformément à la norme ISAD(G) des fichiers électroniques fournis par le service d'archives.
- Phase 2 : transformation des fichiers structurés en ISAD(G) au format XML/EAD suivant la DTD EAD 2002 et élaboration de feuilles de transformation XSL permettant la génération de fichiers PDF et/ou HTML.

Autre scénario possible dans le cas de la conversion en EAD d'une base de données documentaire :

- *Phase 1 : étude par le prestataire de la structure source et de la structure cible (contextualiser les unités descriptives, identifier, pour chaque champ de la base, l'(les) élément(s) / l'(les) attribut(s) équivalent(s) de l'EAD ; déterminer les cas de parfaite équivalence entre la structure source et la*

structure cible, les cas où les informations présentes dans deux champs différents ont vocation à être encodées dans un seul élément EAD ou au contraire où les informations présentes dans un seul champ devront être réparties entre deux éléments EAD, etc.).

- Phase 2 : élaboration par le prestataire d'un programme de conversion puis reprise du résultat de la conversion.
- Phase 3 : conversion des fichiers EAD 2002 en fichiers conformes au profil apeEAD élaboré pour le Portail européen des archives (<http://www.archivesportaleurope.net/>) au moyen de l'outil de conversion fourni dans le cadre du projet (<http://www.apenet.eu/> Rubrique " A propos d'APEnet / Développements ").
- Phase 4 : livraison au service d'archives des fichiers XML EAD 2002 et apeEAD et des feuilles de style par supports amovibles ou *via* un transfert réseau.

Il conviendra de s'assurer de la conservation pérenne des fichiers EAD et des feuilles de style (support amovible pérenne ou intégration dans un système d'information sécurisé dont peuvent disposer les archives ; dans ce dernier cas, le transfert devra se faire par réseau FTP). Les fichiers devront être répartis en trois dossiers : un dossier pour les fichiers en EAD 2002, un dossier pour les feuilles de style et un dernier dossier pour les fichiers apeEAD.

Pour les phases 1 et 2, il est demandé au prestataire de présenter au service d'archives sa méthode de travail en matière de structuration des fichiers conformément à la norme ISAD(G) et de production de fichiers XML.

5.3. Description des différentes phases

Phase 1 : les instruments de recherche antérieurs à 1994, tout en s'approchant de la *Norme générale internationale de description archivistique* ou norme ISAD(G), ne la respectent pas pour autant. Par conséquent, le service d'archives demandera au prestataire certaines opérations de transformation de l'instrument de recherche pour rendre les fichiers électroniques remis par les archives conformes à la norme ISAD(G) (<http://www.ica.org/?lid=10225>).

Pour chaque instrument de recherche, le prestataire proposera au service d'archives des spécifications détaillant les opérations de normalisation à appliquer. Ces spécifications seront examinées et éventuellement amendées par les archives, puis rédigées dans leur version définitive par le prestataire.

La conception des spécifications détaillées par le prestataire est facultative dans la mesure où le service d'archives est susceptible d'élaborer lui-même ces spécifications détaillées.

Exemples :

- ajout des cotes et des dates extrêmes pour les niveaux de titre ne les ayant pas ;

- transformation des “ idem ” en titres de niveau intermédiaire (série organique) pour lesquels le prestataire devra indiquer les cotes extrêmes et les dates extrêmes ;
- suppression des * après la cote ;
- normalisation des intitulés/analyses en en retrouvant puis en en ré-agencant les différentes parties (notamment, objet, action, typologie) avec une ponctuation et dans un ordre conforme aux prescriptions des Archives de France ;
- création d'une description globale pour les articles qui en sont dépourvus à partir des intitulés des dossiers qu'ils contiennent et sur la base de consignes détaillées données par le service d'archives ;
- transformation des dates révolutionnaires, approximatives et littérales sous leur forme numérique (AAAA, MM-AAAA, JJ-MM-AAAA).

(Optionnel)

Il peut être demandé au prestataire d'indexer les noms de personnes, de familles, de collectivités et les noms géographiques (ou de reprendre cette indexation, quand elle existe), en conformité avec le fichier d'autorité des administrations françaises depuis 1800 élaboré par le Service interministériel des Archives de France et l'Association des archivistes français (<https://www.ica-atom.org/aaf/>), ainsi qu'en fonction des règles précisées par la norme AFNOR NF Z 44-060 pour les noms de collectivités², la norme AFNOR NF Z 44-061 pour les noms de personnes³, et la norme AFNOR NF Z 44-081 pour les noms géographiques⁴.

- *Si l'instrument de recherche a déjà fait l'objet d'une indexation, le prestataire devra, pour chaque entrée d'index existante :*
 - o *vérifier, à partir des référentiels appropriés, qu'elle correspond à une forme autorisée ;*
 - o *retrouver éventuellement la forme autorisée correspondante dans les référentiels appropriés et remplacer le terme d'origine par cette forme.*
- *Si l'instrument de recherche n'a pas été indexé, le prestataire devra repérer et indexer en fonction des référentiels qui lui seront fournis les noms de personnes, de familles, de collectivités, les noms géographiques.*
- *Si aucune entrée appropriée n'existe dans les référentiels mentionnés ci-dessus, il reviendra au prestataire de proposer un candidat descripteur, sur la base des règles et consignes fournies par le service d'archives et de remplacer le terme d'origine par ce terme après validation du service.*

(Optionnel)

Il peut également être demandé au prestataire de procéder à une indexation matière. Le service d'archives signale par l'emploi dans l'instrument de recherche d'une mise en forme particulière

² NF Z 44-060 Décembre 1996 - *Documentation - Catalogue d'auteurs et d'anonymes - Forme et structure des vedettes de collectivités-auteurs.*

³ NF Z 44-061 Juin 1986 - *Documentation - Catalogage - Forme et structure des vedettes noms de personnes, des vedettes titres, des rubriques de classement et des titres forgés.*

⁴ NF Z 44-081 Septembre 1993 - *Documentation - Catalogage des documents cartographiques - Forme et structure des vedettes noms géographiques.*

les termes qu'il souhaite voir indexés ; le prestataire recherche le descripteur approprié dans le Thésaurus pour la description et l'indexation des archives locales (<http://data.culture.fr/thesaurus/>) et procède à l'indexation.

Phase 2 : les instruments de recherche restructurés en ISAD(G) et à convertir au format XML le seront selon la DTD EAD 2002. Des feuilles de transformation XSL permettant la génération de fichiers PDF et/ou HTML seront fournies avec les fichiers XML.

La phase 2 débutera après la validation par le service d'archives des fichiers restructurés en ISAD(G). Des prescriptions sont fournies ci-dessous au prestataire concernant l'encodage général des instruments de recherche. Le prestataire devra soumettre aux archives par voie électronique, au fur et à mesure qu'ils se présenteront, les problèmes ou anomalies non résolus.

Les fichiers XML/EAD ainsi produits seront fournis par voie électronique au service d'archives pour validation. Les archives disposeront d'un délai de XX jours à compter de la date de réception des fichiers XML/EAD pour valider ces fichiers.

- Règles générales d'encodage :

Le prestataire respectera les règles générales suivantes lors de l'application de l'EAD à l'instrument de recherche :

- Un fichier XML sera produit pour chaque instrument de recherche.

Dans certains cas, il peut être nécessaire de produire, pour un seul instrument de recherche, plusieurs fichiers XML correspondant aux différentes sous-séries du plan de classement. Dans d'autres cas, il pourra être nécessaire de fusionner plusieurs instruments de recherche pour produire un seul instrument de recherche. Voir plus haut les remarques sur la granularité des instruments de recherche (paragraphe 2.2).

- L'intégralité des informations contenues dans l'instrument de recherche sera convertie, à l'exception des préfaces, avant-propos, remerciements éventuels, tableaux méthodiques de noms de lieux et des mots-matières.

La prise en compte de tel ou tel type d'informations pour l'encodage est à déterminer au cas par cas. Par exemple, il peut être utile d'inclure dans l'encodage les préfaces tenant lieu d'introductions historiques.

- Chaque fichier est un fichier XML valide. L'ensemble du fichier est englobé dans la balise <ead> et comprend les deux grandes balises <eadheader>, qui contient les métadonnées de l'instrument de recherche, et <archdesc> qui inclut la description archivistique. On n'utilisera pas l'élément Préliminaires <frontmatter>.

L'élément Préliminaires <frontmatter> ne figurera plus dans le futur schéma

EAD que la Société des archivistes américains devrait publier fin 2015. Si le contenu d'un élément <frontmatter> est jugé indispensable, il pourra être encodé dans un élément de la balise <scopecontent>.

- Le prestataire évitera au maximum d'utiliser les éléments ou attributs de mise en forme, ou ceux destinés à titrer les données, sauf exception et à la demande du service d'archives. Il est notamment demandé de ne pas utiliser l'attribut @type="titre" dans <unittitle> et de ne pas utiliser de balise <lb/> dans <physdesc>, <repository>, etc.

- Il est demandé au prestataire de renseigner la balise <eadheader> avec les informations relevant des métadonnées bibliographiques à encoder conformément à l'exemple suivant :

```
<ead audience="external">
<eadheader countryencoding="iso3166-1" dateencoding="iso8601"
langencoding="iso639-2b" repositoryencoding="iso15511"
scriptencoding="iso15924" relatedencoding="MARC21">
<eadid identifier="FR-FRAD000_0000_SERIE_00" countrycode="FR"
mainagencycode="FR-FRAD000">FR-FRAD000_0000_SERIE_00</eadid>
<filedesc>
<titlestmt>
<titleproper>Intitulé du fonds (de la série)</titleproper>
<subtitle>Type d'instrument de recherche</subtitle>
<author>Auteur</author>
</titlestmt>
<editionstmt>
<edition>Première édition électronique réalisée à partir de l'édition
imprimée de 19xx</edition>
</editionstmt>
<publicationstmt>
<publisher encodinganalog="260$b">Archives départementales de
XXX</publisher>
<address><addressline>Ville</addressline></address>
<date calendar="gregorian" era="ce" normal="AAAA">AAAA</date>
</publicationstmt>
</filedesc>
<profiledesc>
<creation>Encodage en XML EAD réalisé par XXX pour les archives
départementales de XXX< calendar="gregorian" era="ce" date normal="2013-MM
JJ">jour mois 2013</date></creation>
<langusage>Cet instrument de recherche est
en<language>français</language></langusage>
</profiledesc>
</revisiondesc>
</eadheader>
```

- L'identifiant du fichier comprendra différents éléments séparés par des tirets underscore :

- o FR-FRAD000 : code institutionnel du service d'archives ;

- APE pour les fichiers XML au format apeEAD ;
- code opération (sur quatre chiffres) interne au service d'archives ;
- identifiant de la série correspondant à l'instrument de recherche encodé ;
- numéro d'ordre du fichier

Exemples :

```
<eadid identifier="FR-FRAD018_0008_1M_01" countrycode="FR"
mainagencycode="FR-FRAD000">FRAD018_0008_1M_01</eadid>
```

(fichier XML EAD2002)

```
<eadid identifier="FR-FRAD000_APE-0008_1M_01" countrycode="FR"
mainagencycode="FR-FRAD000">FRAD018_APE_0008_1M_01</eadid>
```

(fichier XML apeEAD)

Dans le profil européen de l'EAD, le contenu de l'élément <eadid> ainsi que les valeurs de ses attributs @identifier et @mainagencycode sont contraints par rapport à la norme ISO 15511 selon laquelle le code d'une institution doit comprendre le code pays et le code de l'institution séparés par un tiret. Pour se conformer à ces exigences, le code institutionnel pour les archives départementales sera donc : FR-FRAD.

- Certains éléments doivent systématiquement être complétés par un ou plusieurs attributs, en plus de ceux précisés ponctuellement dans les spécifications techniques (voir annexe 2).

- Introduction :

- Les éléments composant l'introduction historique et la présentation des fonds seront dans la mesure du possible décrits dans Description archivistique <archdesc>, en les répartissant entre les balises utilisées à ce niveau de description. Si le texte d'origine ne se prête à cette opération qu'au prix de nombreuses modifications, il sera alors décrit non pas dans <archdesc> mais dans Présentation du contenu <scopecontent> avec les autres parties liminaires, sans d'autres balises que celles utilisées pour reproduire les titres et alinéas d'origine (<head> et <p>). Le prestataire devra néanmoins, dans ce cas, en tirer les éléments les plus généraux pour renseigner au minimum, dans l'élément Identification et description <did> de la balise <archdesc>, les éléments <origination> (provenance), <unitid> (cotes extrêmes), <unittitle> (intitulé) et <unitdate> (dates extrêmes).

- Corps de l'instrument de recherche :

- Pour encoder le corps de l'instrument de recherche, on utilisera un ou plusieurs éléments Description des sous-composants <dsc> contenant des éléments composants <c> non numérotés. Chaque élément <c> contiendra un élément Identification et description <did> et ses sous-éléments (<unitid>, <unitdate>, <unittitle> notamment, éventuellement <origination> et <physdesc>) et, le

cas échéant, les autres balises appropriées, notamment <scopecontent>. Les mentions de non communicabilité figurant dans l'instrument de recherche seront reprises dans la balise <accessrestrict>.

- Les titres et sous-titres de l'instrument de recherche seront restitués sous forme d'éléments <c> englobants, dans la balise <unittitle> de l'élément <did>. Le prestataire devra par ailleurs compléter la description de ce composant en ajoutant les éléments <unitid> (cotes extrêmes des éléments englobés) et <unitdate> (dates extrêmes des éléments englobés), qui généralement n'apparaissent pas dans les instruments de recherche papier. Les éléments de numérotation des titres seront en revanche supprimés dans le document encodé.

- Le prestataire donnera une valeur à l'attribut @level de l'élément <archdesc> et de chacun des éléments <c>. Seules les valeurs "fonds", "series", "subseries", "file" et "item" pourront être utilisées. Seuls les emboîtements suivants sont autorisés :

- <c level= "fonds"> peut contenir <c level= "fonds">, <c level= "series">, <c level= "subseries">, <c level= "file"> ;
- <c level= "series"> peut contenir <c level= "series">, <c level= "subseries"> <c level= "file">, <c level= "item"> ;
- <c level= "subseries"> peut contenir <c level= "subseries">, <c level= "file">, <c level= "item"> ;
- <c level= "file"> peut contenir <c level= "item">.

Scénario alternatif : le prestataire ne renseignera que le @level de <archdesc>, obligatoire, et les <c> correspondant aux niveaux les plus élevés et les plus fins du corps de l'instrument de recherche. Dans l'EAD comme dans le profil européen de l'EAD, l'attribut @level des composants <c> n'est obligatoire qu'au plus haut niveau, ensuite il ne l'est pas. Le caractère non obligatoire de @level sauf au niveau le plus haut de la description devrait être maintenu dans la future version de l'EAD.

- Le prestataire donnera une valeur à l'attribut @id de chacun des éléments <c>, selon le modèle suivant :

id="cCOTESANSESPACE" ou id="cCOTESANSESPACE_COTESANSESPACE".

Exemples :

- la cote 1 M 360 sera encodée dans l'élément <c level="..." id="c1M360">
- les cotes extrêmes 1 M 350 - 1 M 360 seront encodées dans l'élément <c level="..." id="c1M350_1M360">

Dans le profil européen, les identifiants @id ne sont pas obligatoires pour les <c>.

- L'élément <unitdate> étant destiné à devenir un point d'accès important (à produire un index des dates permettant de faire des recherches sur les dates des documents), il est jugé nécessaire de saisir, en plus de la forme " lisible " une forme normalisée de la date encodée. Cette forme normalisée doit être systématiquement saisie dans l'attribut @normal de l'élément <unitdate>, en prenant comme référentiel la norme ISO 8601 (voir annexe 3).

Exemples :

```
<unitdate normal="1867/1912">1867-1912</unitdate>  
<unitdate normal="1790-01-01/1790-12-31">1790</unitdate>  
<unitdate normal="1801/1900">XIXe siècle</unitdate>
```

- Si des dates apparaissent sous la forme du calendrier républicain (“ 15 prairial an V ” par exemple), cette forme sera gardée dans l’élément <unitdate> ou <date>, mais elles seront transcrites sous la forme du calendrier grégorien dans l’attribut normal (voir annexe 3).

- La mention s.d. (sans date) sera encodée avec l’élément <unitdate>

- Les dates composées d’une fourchette chronologique et d’une date isolée seront encodées dans deux éléments <unitdate> différents.

Exemple :

```
<did>  
<unitdate normal="1815/1836">1815-1836</unitdate>  
<unitdate normal="1845">1845</unitdate>  
</did>
```

- Généralement, la présentation d’un instrument de recherche est la suivante :

Cote Intitulé Dates extrêmes

qui donnera lieu à l’encodage suivant :

```
<c level="..." id="...">  
<did>  
<unitid>Cote</unitid>  
<unittitle>Intitulé</unittitle>  
<unitdate normal="...">Dates extrêmes</unitdate>  
</did>  
</c>
```

- Si l’intitulé de l’unité documentaire se réduit à une date ou à un intervalle de dates, comme souvent pour une série de dossiers classés dans l’ordre chronologique et dont le contenu est une séquence chronologique, le prestataire saisira la date sous la forme “ lisible ” dans l’élément <unittitle> et sous la forme normalisée dans l’attribut @normal de <unitdate>.

Exemple :

Cote dates extrêmes

qui donnera lieu à l’encodage suivant :

```
<c level="..." id="...">  
<did>  
<unitid>Cote</unitid>  
<unittitle>Dates extrêmes</unittitle>  
<unitdate normal="..."></unitdate>  
</did>  
</c>
```

Cette spécification tient compte des évolutions à venir de l’EAD. Dans le futur

schéma de l'EAD, l'élément <unitdate> ne devrait plus être disponible dans <unititle>.

Les informations relatives à l'importance matérielle et au support des documents décrits devront être saisies soit sous une forme structurée dans les sous-éléments de <physdesc> (<extent>, <dimension>, <physfacet>, <genreform>) soit sous forme textuelle directement dans <physdesc>. L'élément <physdesc> ne pourra pas contenir à la fois des informations textuelles et des sous-éléments.

Exemple :

```
<physdesc>
<extent type="nombre_articles">542 articles</extent><extent
type="metrage">40 mètres linéaires</extent>
</physdesc>
```

```
<physdesc>542 articles, 40 mètres linéaires</physdesc>
```

Cette spécification tient compte des évolutions à venir de l'EAD. Dans le futur schéma de l'EAD, les informations relatives à l'importance matérielle devront être saisies soit sous une forme structurée dans un élément <physdescstructured>, soit sous une forme textuelle non structurée dans un élément <physdesc>.

(Optionnel) :

- Les renvois (renvoi de la description d'une cote à une autre description, par exemple) faits par note en bas de page ou dans le corps du texte de l'instrument de recherche seront restitués en utilisant les éléments de lien interne de l'EAD (attributs @id de l'élément <c> et @target de l'élément <ref>).

- Les autres données inscrites en notes de bas de page seront réintégrées dans le corps de l'instrument de recherche, au sein d'autres balises plus appropriées que la balise <note> (<bioghist>, <custodhist>, <scopecontent>, <physdesc>, <accessrestrict>, etc.).

- L'attribut @encodinganalog sera renseigné dans <unititle>, <unitdate>, <unitid>, <custodhist>, <accessrestrict>, <userrestrict>, etc. pour indiquer les équivalences avec les éléments de la norme ISAD(G) (voir annexe 2 pour les différentes valeurs possibles pour @encodinganalog).

Exemple :

```
<c level="..." id="..." encodinganalog="3.1.4">
<did>
<unitid encodinganalog="3.1.1">Cote</unitid>
```

```

<unittitle encodinganalog="3.1.2">Intitulé</unittitle>
<unitdate encodinganalog="3.1.3" calendar="gregorian" era="ce"
normal="...">Dates</unitdate>
<physdesc encodinganalog="3.1.5">Description physique</physdesc>
<origination encodinganalog="3.2.1">Provenance</origination>
</did>
<accessrestrict encodinganalog="3.4.1"><p>Conditions
d'accès</p></accessrestrict>
<userrestrict encodinganalog="3.4.5"><p>Conditions d'utilisation</p>
</userrestrict>
<otherfindaid encodinganalog="3.4.5"><p>Instrument de recherche associé</p>
</otherfindaid>
</c>

```

- L'indexation sera faite par l'utilisation des balises en <controlaccess> à la fin de chaque unité documentaire. Les différentes balises de <controlaccess> seront utilisées en fonction des typologies des descripteurs : <subject> (mots-matières), <persname> (noms de personnes), <corpname> (noms de collectivité), <famname> (noms de familles), <function> (domaines d'intervention administrative), <occupation> (fonctions exercées par des individus), <genreform> (typologies documentaires).

Cette spécification tient compte des évolutions à venir de l'EAD. Dans le futur schéma de l'EAD, les éléments d'indexation ne devraient plus être disponibles dans <unittitle>, seulement dans le <p> de <scopecontent>.

Pour chaque instrument de recherche, le prestataire proposera au service d'archives des spécifications d'encodage. Ces spécifications seront examinées et éventuellement amendées par les archives, puis rédigées dans leur version définitive par le prestataire.

La conception des spécifications détaillées par le prestataire est facultative dans la mesure où le service d'archives est susceptible d'élaborer lui-même ces spécifications détaillées.

Le prestataire fournira au fur et à mesure de leur réalisation, par messagerie électronique ou protocole FTP, les fichiers encodés au format XML/EAD.

Le service d'archives vérifiera qu'ils sont conformes aux caractéristiques techniques définies avec le prestataire.

Afin de pouvoir exploiter les instruments de recherche encodés en EAD2002 fournis par le prestataire, le service d'archives tient à disposer de feuilles de transformation XSL permettant la génération de fichiers PDF et/ou HTML. Le prestataire développera une feuille de transformation par instrument de recherche encodé, permettant notamment de générer des index et de les faire figurer en fin d'instrument de recherche.

Phase 3 : conversion des fichiers EAD 2002 en fichiers conformes au profil apeEAD élaboré pour le Portail européen des archives (<http://www.archivesportaleurope.net/>) au moyen de l'outil de conversion fourni dans le cadre du projet (<http://www.apenet.eu/> Rubrique " A propos d'APEnet / Développements ").

Phase 4 : livraison au service d'archives des fichiers EAD 2002 et apeEAD et des feuilles de style par supports amovibles ou *via* un transfert réseau.

5.4. Règles pour le nommage des fichiers produits

Les fichiers seront nommés conformément au modèle suivant, préconisé par le ministère de la Culture et de la Communication (département de la recherche, de l'enseignement supérieur et de la technologie) :

- FRAD000 : code normalisé de l'institution désignant le service d'archives ;
- APE pour les fichiers XML au format apeEAD
- code opération (sur quatre chiffres) interne aux archives départementales ;
- identifiant de la série correspondant à l'instrument de recherche encodé ;
- numéro d'ordre de chaque fichier, si besoin, suite numérique sur deux caractères uniquement ;
- EXT : extension déterminant le format physique du fichier (ici XML).

Exemples :

FRAD018_0008_1M_01.xml (fichier XML EAD2002)

FRAD018_APE_0008_1M_01.xml (fichier XML apeEAD)

5.5. Description des modalités de contrôle qualité de l'encodage

Le contrôle consistera à vérifier :

- l'intégralité de la prestation (mise en conformité des instruments de recherche par rapport à la norme ISAD(G), présence des fichiers encodés en EAD 2002 et en apeEAD, ainsi que des feuilles de transformation XSL et des rapports de production) ;
- la cohérence (respect des règles de nommage énoncées au paragraphe 5.4) ;
- en cas de choix d'un support amovible pérenne, la qualité des disques (le prestataire s'appuiera sur les recommandations des Archives de France basées sur les tests du LNE⁵) ;
- l'exactitude de la structuration des instruments de recherche en ISAD(G), en respect des préconisations mises au point conjointement avec le prestataire lors de la phase 1 ;
- la qualité de l'encodage en XML conformément à la DTD EAD 2002 en respect des préconisations présentes dans le CCTP et de celles mises au point lors de la phase 2 ;
- la qualité de la conversion des fichiers EAD 2002 au format apeEAD en conformité avec les

⁵ On pourra consulter les différents rapports de test du LNE commandés par les Archives de France pour les CD, DVD et Blu-Ray à la page suivante : <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/formats-et-supports/>

guides de bonnes pratiques et outils de conversion et de validation disponibles sur le site du projet de portail européen des archives ;

- la qualité des feuilles de style XSL permettant la transformation des fichiers EAD2002 en fichiers PDF et/ou HTML.

Pour les fichiers restructurés en ISAD(G), une relecture complète des données à restructurer ou à transformer sera réalisée, la qualité des fichiers restructurés étant les fichiers sources pour la rétroconversion. Les points suivants feront l'objet d'une attention particulière :

- restructuration de l'introduction en ISAD(G) ;
- mise en œuvre de la description à plusieurs niveaux (hiérarchisation des niveaux de description, non répétition des informations) ;
- homogénéité de la présentation des notices descriptives ;
- cohérence de l'indexation (*si cette prestation a été demandée*).

Pour les fichiers XML, on effectuera :

- une vérification technique :
 - o la validation des fichiers se fera avec deux éditeurs XML et les outils de validation développés pour le projet de portail archivistique européen ; toute erreur (fichier EAD 2002 non valide par rapport à la DTD EAD 2002, fichier apeEAD non valide par rapport au profil apeEAD) entraînera le rejet de la prestation dans son intégralité ;
 - o la complétude du fichier sera contrôlée par rapport à l'instrument de recherche fourni ;
- une relecture complète des données encodées dans les éléments <eadheader>, <archdesc> jusqu'à et à l'exclusion de <dsc> pour tous les instruments de recherche. Si le nombre d'erreurs rencontrées (absence d'un élément ou d'un attribut, contenu de l'élément ou de l'attribut non conforme aux spécifications) est supérieur à 1 pour un instrument de recherche donné, le prestataire devra livrer dans les XX jours la version corrigée du fichier, qui sera à nouveau soumise à la vérification du service d'archives dans les mêmes conditions ;
- un contrôle de la qualité de l'élément <dsc> sur la base d'un échantillonnage au sein de chaque fichier XML. L'échantillon sera composé, d'une part de l'intégralité des données décrivant un sous-fonds, une série ou une sous-série, d'autre part d'une sélection d'éléments <c> de bas niveau choisis de façon aléatoire dans l'instrument de recherche. L'échantillon ainsi formé représentera environ 20 % des pages imprimées de l'instrument de recherche. Si le nombre d'erreurs constatées lors de la lecture de l'échantillon (absence d'un élément ou d'un attribut ; contenu de l'élément ou de l'attribut non conforme aux spécifications ; position de l'élément dans la hiérarchie des composants EAD non conforme aux spécifications) dépasse 1 erreur pour un instrument de recherche donné, le prestataire devra livrer dans les XX jours la version corrigée du fichier, qui sera à nouveau soumise à la vérification des archives dans les mêmes conditions.

Annexe 1 - Exemple de grille descriptive d'instruments de recherche

Série	Sous-série	Auteur Titre Tome année d'édition	format en cm fermé	format en cm ouvert	type d'instrument de recherche	support matériel	nombre de pages d'intro	nombre de pages du corps de l'IR	nombre de pages d'annexes	nombre de pages d'index	nombre de pages pour table des matières et autres	nombre de pages total	nombre d'articles	notes de bas de page (nombre de lignes + détails)	problèmes potentiels	cible numérique ⁶

⁶ Instruments de recherche à maintenir tels quels, à fusionner ou à éclater..

Annexe 2 - Attributs à ajouter à certains éléments

Les attributs marqués en **gras** sont à ajouter obligatoirement.

Élément	Attribut	Contenu de l'attribut
<accessrestrict>	@encodinganalog	3.4.1 [équivalence avec ISAD(G)]
<acqinfo>	@encodinganalog	3.2.4 [équivalence avec ISAD(G)]
<appraisal>	@encodinganalog	3.3.2
<archdesc>	@level	fonds
	@encodinganalog	3.1.4 [équivalence avec ISAD(G)]
	@relatedencoding	ISAD(G)v2
<arrangement>	@encodinganalog	3.3.4 [équivalence avec ISAD(G)]
<author>	@encodinganalog	245\$c [équivalence avec MARC21]
<bibliography>	@encodinganalog	3.5.4 [équivalence avec ISAD(G)]
<bioghist>	@encodinganalog	3.2.2 [équivalence avec ISAD(G)]
<c>	@level	fonds, series, subseries, file, item
	@encodinganalog	3.1.4 [équivalence avec ISAD(G)]
	@id	[cCOTESANSESPACE] ou [cCOTESANSESPACE_COTESANSESPACE]
<corpname>	@normal	[donner la forme normalisée du nom]
	@role	[préciser le rôle joué par la collectivité par rapport aux documents décrits (producteur, auteur, sujet...)]
	@source	[préciser le référentiel d'où provient la forme normalisée]
	@authfilenumber	[identifiant de la notice d'autorité au sein du référentiel concerné]
<custodhist>	@encodinganalog	3.2.3 [équivalence avec ISAD(G)]
<date>	@calendar	gregorian
	@era	ce
	@normal	[date normalisée]
<ead>	@audience	external
<eadheader>	@countryencoding	iso3166-1
	@dateencoding	iso8601
	@langencoding	iso639-2b
	@repositoryencoding	iso15511
	@scriptencoding	iso15924
	@relatedencoding	MARC21
	@countrycode	FR

	@identfier	<i>FR-FRAD000_XXX [Identifiant de l'instrument de recherche]</i>
	@mainagencycode	FR-FRAD000
<famname>	@normal	<i>[donner la forme normalisée du nom]</i>
	@role	<i>[préciser le rôle joué par la famille par rapport aux documents décrits (producteur, auteur, sujet...)]</i>
	@source	<i>[préciser le référentiel d'où provient la forme normalisée]</i>
	@authfilenumber	<i>[identifiant de la notice d'autorité au sein du référentiel concerné]</i>
<function>	@normal	<i>[donner la forme normalisée]</i>
	@source	<i>[préciser le référentiel d'où provient le terme]</i>
	@authfilenumber	<i>[identifiant du terme au sein du référentiel concerné]</i>
<genreform>	@normal	<i>[donner la forme normalisée]</i>
	@source	<i>[préciser le référentiel d'où provient le terme]</i>
	@authfilenumber	<i>[identifiant du terme au sein du référentiel concerné]</i>
<geogname>	@normal	<i>[donner la forme normalisée du nom]</i>
	@source	<i>[préciser le référentiel d'où provient la forme normalisée]</i>
	@authfilenumber	<i>[identifiant de la notice d'autorité au sein du référentiel concerné]</i>
<language> (élément de <langusage)	@langcode	fre
	@scriptcode	Latn
	@encodinganalog	041
<language> (élément de <langmaterial)	@langcode	fre
	@scriptcode	Latn
	@encodinganalog	3.4.3 [équivalence avec ISAD(G)]
<occupation>	@normal	<i>[donner la forme normalisée]</i>
	@source	<i>[préciser le référentiel d'où provient la forme normalisée]</i>
	@authfilenumber	<i>[identifiant du terme au sein du référentiel concerné]</i>

<origination>	@encodinganalog	3.2.1 [équivalence avec ISAD(G)]
<persname>	@normal	[donner la forme normalisée du nom]
	@role	[préciser le rôle joué par la personne par rapport aux documents décrits (producteur, auteur, sujet...)]
	@source	[préciser le référentiel d'où provient la forme normalisée]
	@authfilenumber	[identifiant de la notice d'autorité au sein du référentiel concerné]
<physdesc>	@encodinganalog	3.1.5 [équivalence avec ISAD(G)]
<publisher>	@encodinganalog	260\$b
<relatedmaterial>	@encodinganalog	3.5.3 [équivalence avec ISAD(G)]
<scopecontent>	@encodinganalog	summary
<subject>	@normal	[donner la forme normalisée]
	@source	[préciser le référentiel d'où provient le terme]
	@authfilenumber	[identifiant du terme au sein du référentiel concerné]
<unitdate>	@calendar	gregorian
	@era	ce
	@normal	[dates sous leur forme normalisée selon la norme ISO 8601]
	@encodinganalog	3.1.3 [équivalence avec ISAD(G)]
<unitid>	@encodinganalog	3.1.1 [équivalence avec ISAD(G)]
	@type	file_reference [pour la cote d'archives] call_number [pour la cote de communication]
<unittitle>	@encodinganalog	3.1.2 [équivalence avec ISAD(G)]
<userrestrict>	@encodinganalog	3.4.2 [équivalence avec ISAD(G)]

Annexe 3 – Valeurs de l’attribut @normal de l’élément <unitdate>

Forme rédigée de la date	Valeur de l’attribut normal	Commentaire
1822	1822	
XVIIIe siècle	1701/1800	
1785-1836	1785/1836	
1792-mars 1800	1792/1800-03	
2 octobre 1802-3 décembre 1804	1802-10-02/1804-12-03	
an V	1796/1797	Convertir dans l’attribut normal les dates révolutionnaires en intervalles de dates (voir tableau de concordance ci-dessous)
1769-anVIII	1769/1800	
an XI - 1942	1802/1942	
XVIIe-XVIIIe siècle	1601/1800	
1819 - ...	1819/2099	
1758 - ... (série ouverte)	1758/2099	
Fin XVIIe siècle- ...	1670/2099	Donner une fourchette de 30 ans.
XVe siècle-1788 (sous série A1) et 1791-1812 (série B)	<unitdate normal="1401/1788">XVe siècle-1788 (sous série A1)</unitdate> <unitdate normal="1791/1812">1791-1812 (série B)</unitdate>	Rupture entre deux dates ou intervalles de dates : employer deux éléments <unitdate>.
Origines-1816	0000/1816	
Non précisée	Pas d’attribut normal.	

Concordance calendrier révolutionnaire / calendrier grégorien

Calendrier républicain : de l'an II à l'an XIV

Mois républicains	an II 1793-1794	an III 1794-1795	an IV 1795-1796	an V 1796-1797	an VI 1797-1798	an VII 1798-1799
1er Vendémiaire	22 sept. 1793	22 sept. 1794	23 sept. 1795	22 sept. 1796	22 sept. 1797	22 sept. 1798
15 Vendémiaire	6 oct. 1793	6 oct. 1794	7 oct. 1795	6 oct. 1796	6 oct. 1797	6 oct. 1798
1er Brumaire	22 oct. 1793	22 oct. 1794	23 oct. 1795	22 oct. 1796	22 oct. 1797	22 oct. 1798
15 Brumaire	5 nov. 1793	5 nov. 1794	6 nov. 1795	5 nov. 1796	5 nov. 1797	5 nov. 1798
1er Frimaire	21 nov. 1793	21 nov. 1794	22 nov. 1795	21 nov. 1796	21 nov. 1797	21 nov. 1798
15 Frimaire	5 déc. 1793	5 déc. 1794	6 déc. 1795	5 déc. 1796	5 déc. 1797	5 déc. 1798
1er Nivôse	21 déc. 1793	21 déc. 1794	22 déc. 1795	21 déc. 1796	21 déc. 1797	21 déc. 1798
15 Nivôse	4 janv. 1794	4 janv. 1795	5 janv. 1796	4 janv. 1797	4 janv. 1798	4 janv. 1799
1er Pluviôse	20 janv. 1794	20 janv. 1795	21 janv. 1796	20 janv. 1797	20 janv. 1798	20 janv. 1799
15 Pluviôse	3 févr. 1794	3 févr. 1795	4 févr. 1796	3 févr. 1797	3 févr. 1798	3 févr. 1799
1er Ventôse	19 févr. 1794	19 févr. 1795	20 févr. 1796	19 févr. 1797	19 févr. 1798	19 févr. 1799
15 Ventôse	5 mars 1794	5 mars 1795	5 mars 1796	5 mars 1797	5 mars 1798	5 mars 1799
1er Germinal	21 mars 1794	21 mars 1795	21 mars 1796	21 mars 1797	21 mars 1798	21 mars 1799
15 Germinal	4 avril 1794	4 avril 1795	4 avril 1796	4 avril 1797	4 avril 1798	4 avril 1799
1er Floréal	20 avril 1794	20 avril 1795	20 avril 1796	20 avril 1797	20 avril 1798	20 avril 1799
15 Floréal	4 mai 1794	4 mai 1795	4 mai 1796	4 mai 1797	4 mai 1798	4 mai 1799
1er Prairial	20 mai 1794	20 mai 1795	20 mai 1796	20 mai 1797	20 mai 1798	20 mai 1799
15 Prairial	3 juin 1794	3 juin 1795	3 juin 1796	3 juin 1797	3 juin 1798	3 juin 1799
1er Messidor	19 juin 1794	19 juin 1795	19 juin 1796	19 juin 1797	19 juin 1798	19 juin 1799
15 Messidor	3 juill. 1794	3 juill. 1793	3 juill. 1796	3 juill. 1797	3 juill. 1798	3 juill. 1799
1er Thermidor	19 juill. 1794	19 juill. 1795	19 juill. 1796	19 juill. 1797	19 juill. 1798	19 juill. 1799
15 Thermidor	2 août 1794	2 août 1795	2 août 1796	2 août 1797	2 août 1798	2 août 1799
1er Fructidor	18 août 1794	18 août 1795	18 août 1796	18 août 1797	18 août 1798	18 août 1799

15 Fructidor	1er sept. 1794	1er sept. 1795	1er sept. 1796	1er sept. 1797	1er sept. 1798	1er sept. 1799
5ème jour complémentaire	21 sept. 1794	21 sept. 1795	21 sept. 1796	21 sept. 1797	21 sept. 1798	21 sept. 1799
6ème jour complémentaire		22 sept. 1795				22 sept. 1799

Calendrier républicain : de l'an VIII à l'an XIV

Mois républicains	an VIII 1799-1800	an IX 1800-1801	an X 1801-1802	an XI 1802-1803	an XII 1803-1804	an XIII 1804-1805	an XIV 1805
1er Vendémiaire	23 sept. 1799	23 sept. 1800	23 sept. 1801	23 sept. 1802	24 sept. 1803	1804 23 sept. 7 oct. 1804 23 oct. 1804 6 nov. 1804	1805 23 sept. 7 oct. 1805 23 oct. 1805 6 nov. 1805
15 Vendémiaire	7 oct. 1799	7 oct. 1800	7 oct. 1801	7 oct. 1802	8 oct. 1803		
1er Brumaire	23 oct. 1799	23 oct. 1800	23 oct. 1801	23 oct. 1802	24 oct. 1803		
15 Brumaire	6 nov. 1799	6 nov. 1800	6 nov. 1801	6 nov. 1802	7 nov. 1803		
1er Frimaire	22 nov. 1799	22 nov. 1800	22 nov. 1801	22 nov. 1802	23 nov. 1803	22 nov. 1804	22 nov. 1805
15 Frimaire	6 déc. 1799	6 déc. 1800	6 déc. 1801	6 déc. 1802	7 déc. 1803	6 déc. 1804	6 déc. 1805
1er Nivôse	22 déc. 1799	22 déc. 1800	22 déc. 1801	22 déc. 1802	23 déc. 1803	22 déc. 1804	22 déc. 1805
15 Nivôse	5 janv. 1800	5 janv. 1801	5 janv. 1802	5 janv. 1803	6 janv. 1804	5 janv. 1805	
1er Pluviôse	21 janv. 1800	21 janv. 1801	21 janv. 1802	21 janv. 1803	22 janv. 1804	1805 21 janv. 4 févr. 1805 20 févr. 6 mars 1805	
15 Pluviôse	4 févr. 1800	4 févr. 1801	4 févr. 1802	4 févr. 1803	5 févr. 1804		
1er Ventôse	20 févr. 1800	20 févr. 1801	20 févr. 1802	20 févr. 1803	21 févr. 1804		
15 Ventôse	6 mars 1800	6 mars 1801	6 mars 1802	6 mars 1803	6 mars 1804		
1er Germinal	22 mars 1800	22 mars 1801	22 mars 1802	22 mars 1803	22 mars 1804	1805 22 mars 5 avril 1805 21 avril 5 mai 1805	
15 Germinal	5 avril 1800	5 avril 1801	5 avril 1802	5 avril 1803	5 avril 1804		
1er Floréal	21 avril 1800	21 avril 1801	21 avril 1802	21 avril 1803	21 avril 1804		
15 Floréal	5 mai 1800	5 mai 1801	5 mai 1802	5 mai 1803	5 mai 1804		
1er Prairial	21 mai 1800	21 mai 1801	21 mai 1802	21 mai 1803	21 mai 1804	21 mai 1805	
15 Prairial	4 juin 1800	4 juin 1801	4 juin 1802	4 juin 1803	4 juin 1804	4 juin 1805	
1er Messidor	20 juin 1800	20 juin 1801	20 juin 1802	20 juin 1803	20 juin 1804	20 juin 1805	
15 Messidor	4 juill. 1800	4 juill. 1801	4 juill. 1802	4 juill. 1803	4 juill. 1804	4 juill. 1805	
1er Thermidor	20 juill. 1800	20 juill. 1801	20 juill. 1802	20 juill. 1803	20 juill. 1804	1805 20 juill.	
15 Thermidor	3 août 1800	3 août 1801	3 août 1802	3 août 1803	3 août 1804		

1er Fructidor 15 Fructidor	19 août 1800 2 sept. 1800	19 août 1801 2 sept. 1801	19 août 1802 2 sept. 1802	19 août 1803 2 sept. 1803	19 août 1804 2 sept. 1804	3 août 1805 19 août 1805 2 sept. 1805	
5ème jour complément. 6ème jour complément.	22 sept. 1800	22 sept. 1801	22 sept. 1802	22 sept. 1803 23 sept. 1803	22 sept. 1804	22 sept. 1805	