



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
MINISTÈRE DE L'ÉCOLOGIE, DE L'ÉNERGIE, DU DÉVELOPPEMENT DURABLE
ET DE LA MER

MUSEUM NATIONAL D'HISTOIRE NATURELLE
SERVICE DES MARCHES
57, RUE CUVIER
75231 PARIS CEDEX 05

Tél : 01 40 79 81 31
Fax : 01 40 79 56 10
marches@mnhn.fr

MARCHE N°10M17

PRESTATIONS DE NUMÉRISATION DE DOCUMENTS DES COLLECTIONS
DE LA DIRECTION DES BIBLIOTHÈQUES ET DE LA DOCUMENTATION
DU MUSEUM NATIONAL D'HISTOIRE NATURELLE

DTP

Le présent comporte 19 pages numérotées de 1 à 19

TABLE DES MATIERES

ARTICLE 1 PERIMETRE DU MARCHÉ :	3
LES PRESTATIONS DU PRESENT MARCHÉ COMPRENNENT :	3
ARTICLE 2 MISE A DISPOSITION DES DOCUMENTS.....	4
ARTICLE 3 - NUMERISATION	6
3.1 Définition des prestations	6
3.2 Organisation des répertoires et des fichiers	8
1. Nommage des répertoires et des fichiers.....	8
1.1. Nom du répertoire racine :.....	8
1.2. Nom du répertoire des formats de fichier :.....	9
1.3. Nom du Répertoire de l'unité documentaire (monographie ou fascicule d'un titre de périodique) :	9
1.4. Nom des fichiers :.....	9
2. Nom du répertoire et des fichiers des tables de matières et/ou d'index 10	
3. Arborescence des répertoires.....	10
ARTICLE 4 - RECONNAISSANCE OPTIQUE DES CARACTERES.....	11
ARTICLE 5 - PRESTATIONS D'INDEXATION :SAISIE OU CONVERSION DES TABLES DES MATIERES ET/OU D'INDEX	12
1. Définition	12
2. Conditions d'exécution	13
5.1 Règles particulières de saisie des tables des matières (fichiers xls).....	13
5.2 Règles particulières de saisie des tables d'index (fichiers xml).....	14
5.3 Nommage du répertoire et des fichiers des tables de matières et/ou d'index (voir aussi 1.3.3).	15
1. Nom du répertoire de la table des matières et/ou des index	15
2. Nom des fichiers des tables des matières et/ou des index :.....	15
ARTICLE 6 - PRESTATIONS DE TRANSPORT, DE CONSERVATION ET DE STOCKAGE	16
ARTICLE 7 SUPPORTS DE LIVRAISON.....	16
ARTICLE 8 MODALITES DE CONTROLE PAR LE MNHN	17
8.1 Phase test	17
8.2 Contrôle qualité : procédure	17
a) Erreurs de numérisation.....	18
b) Qualité de la reconnaissance optique de caractères.....	19
c) Erreurs dans les TDM	19

Article 1 Périmètre du marché :

Les prestations du présent marché comprennent :

- des prestations de numérisations,
- des prestations d'indexation,
- prestations de stockage, de conservation et transports des documents originaux / prestations de stockage, de conservation des fichiers numériques
- des prestations de reconnaissance optique de caractères

Le présent cahier des charges décrit les collections à traiter, les prestations de numérisation et d'indexation à effectuer ainsi que les modes de préservation requis lors du transport, du stockage et de la numérisation.

La Bibliothèque remettra au titulaire les documents à numériser accompagnés d'un descriptif de chaque unité documentaire (état matériel, illustrations, exceptions..).

Les documents sont regroupés en fonction de caractéristiques communes, physiques ou de traitement. La préparation de chaque ensemble de documents incombe à la bibliothèque.

Dans le cas présent, 10 ensembles de documents ont été définis

Ensemble de documents	Partie de la collection	Contenu	Spécification
A1	<i>Archives Muséum série 6</i> <i>Archives Muséum série 7</i>	Vol.1-5, 8, 10-13, 16-19 Vol. 1-5, 7-8, 10	massicotables
A2	<i>Archives Muséum série 7</i>	Vol.6 et 9	Volumes avec pages couleur et calques
A3	<i>Archives Muséum série 6,</i> collection pour la communication	Vol. 6,7, 9, 14, 15	exemplaires pour la communication = non massicotables
B1	<i>Bulletin</i>	1895-1906	Bulletins reliés en volumes Massicotables
B2	<i>Bulletin</i>	1907-1969	Bulletins séparés en fascicules massicotables
B3	<i>Bulletin</i>	Zoo n.71,244 + suppl. 1977 + Zoo T. 2 1980	Exemplaires pour la communication = non massicotables

		n° 2 et 4 + Zoo T.6 1984 n°4 + Zoo T.7 1985 n°4 + Zoo T.10 1988 n°1 + Bot. T.8 1986 n°1 + Bot. T. 14 1992 n° 3- 4 + Sc. de la Terre T.1 1979 n°3 + Sc. de la Terre T.8 1986 n°2 + suppl. T. 10 1988	
B4	<i>Bulletin</i>	1970-1978	Bulletins par fascicules massicotables
B5	<i>Bulletin</i>	1979-1996	Bulletins par fascicules massicotables, nouvelle série
B6	<i>Bulletin</i>	1997-1999	Exemplaires pour la communication = non massicotables
M1	<i>Mémoire</i>	1936-1950	Non massicotable

Cette liste n'est pas exhaustive.

Article 2 Mise à disposition des documents

1. Conditions de mise à disposition

Dans un premier temps, la DBD du Muséum fournit au titulaire un ensemble de documents en annexe au Descriptif Technique des Prestations, dont un document Excel décrivant globalement chaque ensemble de documents à numériser.

Ce document est intitulé *description de l'ensemble des documents 2010*. La volumétrie et les documents sont susceptibles d'évoluer en fonction des offres. L'onglet « Cumul ensemble de documents A1 à M1 » donne les totaux de documents à traiter.

Dans un second temps, dans le cadre de l'exécution des prestations du présent marché et après notification, La Bibliothèque remettra au titulaire les documents à numériser, accompagnés d'un descriptif de chaque unité documentaire.

Ce second descriptif indiquera, pour chaque document concerné par les opérations de numérisation :

- le volume des pages à traiter,
- le nombre de planches,
- le nombre de couvertures et de pages couleur,
- toutes les anomalies de pagination et les défauts physiques des objets à numériser, détectés par la Bibliothèque du Muséum.
- Les tables à traiter.

2. Conditions de refus d'un document original par le titulaire

Si un document est remis au titulaire et que ce dernier juge le document non-numérisable (selon les critères définis dans le présent CCP) du fait d'une quelconque de ses caractéristiques physiques, le titulaire signifiera à la Bibliothèque le motif précis de son rejet. Dans chacun de ces cas, la Bibliothèque essaiera de faire parvenir un autre exemplaire au titulaire. Si non, elle statuera sur la suite à donner au document.

Exemples de défauts (liste non-limitative) :

- pagination incohérente (exception faite de la simple inversion de deux pages)
- ouvrages incomplets
- anomalies graves de présentation
- marges étroites < 1 cm (papier uniquement)
- reliures trop serrées
- ouvrages maculés
- détérioration non-signalée
- table des matières unique pour plusieurs volumes, hors cas signalé ;
- partie illisible - en ce cas, le recours au document papier original est possible

3. Contrôle du titulaire

Durant la période où le titulaire a la garde des documents confiés par La DBD du Muséum, obligation lui est faite de respecter toutes les précautions de préservation requises lors du retrait des documents.

Le titulaire communiquera un rapport complet des traitements effectués (problèmes rencontrés, quantités traitées, organisation et déroulement des opérations, etc.)

Il est du devoir titulaire de mettre en place un processus de contrôle de sa production qui lui permette de satisfaire aux exigences édictées dans le présent cahier des charges.

Ce contrôle concerne :

- la qualité de la saisie et de la structuration ;
- la validité des liens ;
- la qualité d'écriture sur disque

- la qualité des images

Après validation des opérations, le titulaire doit :

- soit éliminer les documents papiers si la collection a été désignée comme pouvant être détruite,
- soit, dans le cas contraire (exemplaires de consultation), restituer les documents originaux à la Bibliothèque, dans les mêmes conditions que celles prévalant lors de leur remise au prestataire. Le prestataire devra assurer une constitution de l'ensemble de documents « retour » à l'identique des ensembles de documents « départ ». Les documents ayant subi une dégradation quelconque ou ayant nécessité un traitement particulier seront signalés dans la liste de pointage des documents restitués.

Article 3 - Numérisation

3.1 Définition des prestations

Dans le cadre de l'exécution des prestations du présent marché le titulaire s'engage à :

- numériser les documents originaux. Selon les cas explicitement communiqués par la bibliothèque, cette numérisation pourra se faire avec massicotage ou bien à livre ouvert sans massicotage ni démembrement ; dans certains ensembles de documents, la question du découpage sera précisée dans le bon de commande. A défaut, le découpage sera laissé à l'appréciation du prestataire. En tout état de cause, celui-ci ne devra pas entraîner un surcoût.
- à compresser les images obtenues selon les recommandations qui suivent,
- à les convertir dans les formats souhaités,
- à nommer les fichiers et les répertoires selon les règles fixées par la bibliothèque.

Le titulaire doit être à même de fournir la meilleure qualité de fichier compte tenu de l'usage prévu (ratio qualité d'image / taille de fichier, etc.)

L'utilisateur final devra pouvoir accéder aux documents, afficher et imprimer les pages choisies de façon conviviale et avec une qualité suffisante pour répondre à ses besoins.

La qualité générale de numérisation est liée d'une part à l'état des documents traités mais également à la qualité intrinsèque des moyens de production mis en œuvre par le titulaire.

Le prestataire optimisera la qualité de la restitution ; il ajustera automatiquement le contraste, corrigera les alignements si nécessaire.

Le titulaire s'engage à respecter les normes de corrections graphiques (NF Z 42-013) concernant entre autres le redressement des pages et des éventuelles courbures du texte, le cadrage, le nettoyage des bords.

Du fait de la présence possible de planches en niveau de gris ou couleur dans un même document, le titulaire doit prévoir un traitement mixte NB/gris/couleur, si le cas est signalé dans les annexes.

L'ergonomie et le confort de lecture seront préservés de façon générale (ex : réorientation des pages à l'italienne). En cas de disposition mixte (texte et illustrations par exemple), le sens privilégié sera celui du texte.

Les cas particuliers (présence de dépliants imposant un changement de prise de vue, présence de couvertures cartonnées ou de calques imposant une manipulation) sont signalés ;

- les couvertures sont numérisées ;
- les calques ne sont pas numérisés, sauf s'ils sont légendés ou apportent des informations supplémentaires.

Les recommandations précisées dans les annexes pourront s'appliquer aux particularités d'un document :

- taille des pages,
- optimisation de la résolution,
- spécifications de cadrage,
- rotation de page
- types de tables à saisir ou à convertir

Les fichiers des documents numérisés seront fournis en trois formats distincts pour l'archivage et la diffusion. Les trois formats sont les suivants :

• TIFF 6.0 monopage (une image par page) :

Le format d'archivage TIFF répond à des règles plus précises pour la conservation à long terme. Ce format est utilisé pour l'archivage à long terme.

Résolution demandée :

- Pages de texte : 256 niveaux de gris (8 bits) 600 dpi minimum.
- Planches et figures en niveaux de gris : 256 niveaux de gris 600 dpi minimum
- Planches en couleur : couleur 24 bits à 400 dpi minimum.

Les en-têtes comporteront les (méta)données, au minimum générées automatiquement (informations permettant d'identifier les images, taille d'image, compression...)

• JPG : monopage :

Ce format est utilisé pour la diffusion, il doit donc être un compromis entre qualité de la restitution et poids du fichier.

Résolution demandée : 150 dpi minimum

Une ligne sera automatiquement insérée en bas de page pour identifier l'origine du document (MNHN)

• PDF : monopage (en PDF 1.4 optimisé.)

• PDF : multipage contenant l'ensemble du document (en PDF 1.4 optimisé).

Les formats PDF seront utilisés pour le téléchargement et/ou l'impression des documents.

3.2 Organisation des répertoires et des fichiers

1. Nomage des répertoires et des fichiers

Les règles suivantes sont des règles générales données à titre d'exemple. L'architecture définitive des répertoires et des fichiers pourra être revue avec la Bibliothèque et sous réserve de son accord. Un fichier transmis après notification précisera les noms de répertoires à utiliser.

Sauf cas particulier précisé dans le fichier de description, tous les traitements doivent se faire au niveau du fascicule (chacune des parties détachées d'un ouvrage ou volume publié par livraisons, comportant en général une numérotation et/ou une date de parution uniques)

Remarque : La syntaxe et la longueur des noms devra toujours être la même. Si une des numérotations est absente, remplacer les caractères manquants par des 0.

Exemple : BMAZO_S000_1972_T000_N201_JPG

⇔ Répertoire du *Bulletin du MNHN*, 1972, numéro 201 (ici, numéros de série et de tome inexistant).

1.1. Nom du répertoire racine :

XXXXX (Identifiant de 5 caractères) : Le nom du répertoire racine comprend un champ de 5 caractères identifiant l'ouvrage ou le périodique, par exemple sa cote ou son titre. Les caractères autres qu'alphanumériques (sauf _) et les espaces sont proscrits.

La Bibliothèque fournira après notification un tableau de correspondance entre les titres et les identifiants, à affecter aux noms de répertoires et de fichiers.

Exemple : Titre = *Bulletin du Muséum National d'Histoire Naturelle, section A, Zoologie*
ID = BMAZO

1.2. Nom du répertoire des formats de fichier :

Trois formats de fichiers étant demandés dans trois répertoires distincts (un par format), on donnera à chacun le même nom que le répertoire racine suivi de l'extension correspondant au format, soit : XXXXX (Identifiant de 5 caractères)_TIF
XXXXX_JPG
XXXXX_PDF (avec monopage et multipage)

Exemple : Titre = *Bulletin du Muséum National d'Histoire Naturelle, section A, Zoologie*
ID = BMAZO_PDF => Répertoire contenant tous les fascicules du BMAZO en PDF.

1.3. Nom du Répertoire de l'unité documentaire (monographie ou fascicule d'un titre de périodique) :

⇔ XXXXX_SXXX_XXXX_TXXX_NXXX : Identifiant_série_année_tome_numéro_JPG ou _TIFF ou _PDF (25 caractères)

L'extension à ajouter au nom du répertoire correspond au format des fichiers qu'il contient.

Le caractère « _ » est utilisé comme séparateur.

XXXXX : champs alphanumériques représentant l'identifiant de l'unité documentaire (le même que l'identifiant de la racine).

SXXX : champ de numérotation de la série (3 chiffres maximum précédés de S)

XXXX : champ de l'année de publication. Dans le cas d'une publication sur une période, prendre la 1^{ère} année de la fourchette (1898-1899 deviendra 1899).

TXXX : champ de tomain ou volumain (3 chiffres maximum précédés de T)

NXXX : champ de numérotation de fascicule (3 chiffres maximum précédés de N))

Exemple : BMAZO_S003_1972_T011_N201_PDF
⇔ Répertoire du *Bulletin du MNHN*, 3^e série, 1972, Tome 11, numéro 201 contenant les fichiers en PDF.

1.4. Nom des fichiers :

Il s'agit de fichiers TIFF ou JPG ou PDF

Les noms des fichiers obéissent à la même syntaxe que les noms des répertoires de fascicule, à laquelle s'ajoute : la numérotation de fichier (4 chiffres maximum suivant le rang de numérisation) et l'extension du format.

XXXXX_SXXX_XXXX_TXXX_NXXX_XXXX.tiff ⇔ page tiff

XXXXX_SXXX_XXXX_TXXX_NXXX_XXXX.jpg ⇔ page jpg
XXXXX_SXXX_XXXX_TXXX_NXXX_XXXX.pdf ⇔ page pdf

Remarque : dans le cas du *PDF multipages*, qui correspond à une unité documentaire (le fascicule), il n'y a pas de numéro de page. Le nom du PDF multipage sera le même que le nom du répertoire de fascicule dans lequel il se trouve.

XXXXX_SXXX_XXXX_TXXX_NXXX.pdf ⇔ pdf multipage du fascicule.

Exemple 1 : BMAZO_S003_1972_T011_N201_0025.tiff
⇔ *Bulletin du MNHN*, 3^e série, 1972, Tome 11, numéro 201, vue n°25

Exemple 2 : BMAZO_S003_1972_T011_N201.pdf
⇔ Fichier PDF multipage du fascicule *Bulletin du MNHN*, 3^e série, 1972, Tome 11, numéro 201.

2. Nom du répertoire et des fichiers des tables de matières et/ou d'index

Voir 5.3.

3. Arborescence des répertoires

On aboutit suivant ce plan de nommage à une structure de données de la forme suivante (qui pourra éventuellement être revue avec la Bibliothèque sous réserve de son accord).

Par exemple :

BMAZO	<i>Niveau 1</i> Nom du répertoire contenant toutes les unités documentaires du titre
BMAZO_JPG	<i>Niveau 2</i> Nom du répertoire des JPG rangé dans le répertoire racine
BMAZO_S003_1972_T011_N201_JPG	<i>Niveau 3</i> Nom du répertoire d'une unité documentaire, rangé dans le répertoire de format correspondant

BMAZO_S003_1972_T011_N201_0012.jpg	<i>Niveau 4</i> Nom du fichier correspondant à une page électronique de cette unité documentaire
BMAZO_TDM	<i>Niveau 2</i> Nom du répertoire des fichiers TDM et index
BMAZO_S003_1972_T011_N201_tdm.xls BMAZO_S003_1972_T011_N201_ind.xml	<i>Niveau 3 ou 4</i> Noms du fichier Excel de la table des matières et d'un index pour une unité documentaire des <i>Bulletins</i> , section A, Zoologie.

Article 4 - Reconnaissance optique des caractères

Les pages de texte des documents numérisés feront l'objet d'un traitement, de façon à permettre la recherche sur le texte intégral des fichiers de diffusion, grâce au mode pdf texte sous image.

Le niveau de qualité attendu est celui d'un OCR non corrigé mais le soumissionnaire a toute latitude pour proposer une autre méthode de restitution d'un fichier texte que l'OCR.

Le soumissionnaire indiquera la méthode retenue, les moyens mis en œuvre pour ce traitement ainsi que le niveau de qualité estimé (nombre d'erreurs moyen par mille de caractères).

Le résultat de ce traitement sera également fourni sous le PDF monopage prévu ci-dessus

Article 5 - Prestations d'indexation : Saisie ou conversion des tables des matières et/ou d'index

1. Définition

Certains documents possèdent des tables des matières ou un sommaire, ou des index se rapportant aux illustrations, auteurs, etc. Dans le document imprimé, l'intitulé « table des matières » peut figurer en tête ou en titre d'une table d'index.

- Sont considérées comme **tables des matières (TDM)** les pages dont les libellés renseignent une pagination qui suit le déroulement séquentiel du document du début à la fin. Dans ce groupe, les sommaires sont des formes abrégées des tables des matières.
- Sont considérées comme **tables d'index (IND)** les pages dont les libellés renseignent une pagination qui ne suit pas le déroulement séquentiel de l'ouvrage ou du périodique mais un autre ordre, le plus souvent alphabétique (index d'auteurs, de titres ou de sujets), voire suivant l'ordre alphabétique de rubriques thématiques (tables méthodiques). En outre une entrée d'index peut renvoyer à plusieurs pages.

Remarque : Le titulaire pourra s'appuyer sur les définitions données plus haut et l'aide de la bibliothèque pour faire la distinction entre les deux types de table.

Dans les deux cas, les entrées doivent renvoyer vers une page de l'ouvrage ou du fascicule. Les tables des matières et d'index permettent d'accéder aux pages du document numérisé.

Les tables seront saisies ou converties dans deux fichiers :

- L'un pour les tables des *matières* (enchaînement séquentiel des pages), au format Excel ;
- l'autre pour les tables d'*index* (alphabétiques), au format xml.

Les prestations concernent la saisie de tables des matières et d'index à partir de documents originaux ou de fichiers images. La teneur des tables et des index est donnée pour chaque périodique dans un document annexe de description.

Il y aura un seul fichier par unité documentaire (fascicule ou monographie).

Pour chaque ensemble de documents, la bibliothèque indiquera les tables et index à traiter et signalera les cas particuliers.

Si un document ne comporte aucune table, le prestataire devra en créer au moins une dans le format XLS (ces cas sont signalés et quantifiés en annexe) :

- soit en créant un sommaire à partir du document,

- soit en utilisant les documents annexes mis à sa disposition (photocopies de tables des matières, couvertures des fascicules...)

Remarque : Cette demande ne concerne que les tables des matières absentes pour chaque unité documentaire et non les tables d'index absentes.

Le prestataire peut proposer une autre méthode de restitution d'un fichier texte que la saisie si celle-ci offre une qualité identique pour un coût égal ou inférieur. Le titulaire prendra toute disposition pour vérifier la validité de la saisie.

2. Conditions d'exécution

L'exécution des prestations du présent ensemble de documents implique l'organisation et la mise en œuvre d'une succession d'opérations distinctes concourant à la réalisation de la saisie :

- saisie des données en Unicode UTF-8.
- intitulé des tables : chaque table aura impérativement pour titre les références bibliographiques précises (titre du périodique, numérotation de fascicule, date...)
- positionnement de liens hypertexte ;
- contrôle qualité et contrôle final de production ;
- enregistrement sur le disque dur avec le document numérisé ;
- livraison des documents et des supports d'enregistrement.

5.1 Règles particulières de saisie des tables des matières (fichiers xls)

Les informations de la table des matières seront saisies dans un fichier Excel globalement prévu pour une description plus complète de l'unité documentaire.

Ce fichier, à remplir par le titulaire, indiquera les informations fournies par les tables ou par les documents mais aussi la description bibliographique de l'unité, la correspondance entre les pages logiques et les fichiers, etc. Il y aura 1 fichier excel par fascicule.

Ce fichier comportera ainsi au minimum les informations suivantes, réparties en colonnes :

- numéro de la pagination originale de l'imprimé telle qu'elle se présente dans le document
- nom du fichier numérique correspondant à cette page
- type de la page (ex : couverture, début d'article, illustration, page blanche...)
- Titre de l'article (ou de l'illustration si légende)
- Noms et prénoms des auteurs (ou de l'illustrateur si légende).

Un modèle plus complet est fourni en annexe. Le tableau modèle définitif ainsi que la nomenclature des types de pages à utiliser seront joints après notification.

Cas particulier des illustrations hors texte : les feuilles de planches, les dépliants, les inserts :

Lorsqu'une image n'est pas foliotée dans la pagination de l'ouvrage, il arrive en général que la table des illustrations papier renvoie à la page précédent la feuille de planche ou parfois sur une autre page : le lien n'est donc pas valide. Il appartiendra au prestataire de vérifier sur quelle image pointent les entrées de la table des illustrations (ou figures) et d'indiquer ce numéro d'image dans le lien.

5.2 Règles particulières de saisie des tables d'index (fichiers xml)

Chaque table d'index choisie par la Bibliothèque est reconstituée dans un fichier XML identifié par son en-tête :

```
<?XML VERSION="1.0" ENCODING="UTF-8"?>
<?XML-STYLESHEET TYPE="TEXT/XSL" HREF="DIV_METAFICHIERS-XML.XSL"?>
<METAFILE TYPE="INDEX" COVERAGE="LOCATION"
XMLNS:XSI="HTTP://WWW.W3.ORG/2001/XMLSCHEMA-INSTANCE"
XSI:NONAMESPACESCHEMALOCATION="METAFILE.XSD">
  <CONTROL>
    <MAINENTRYDEPLOY>TRUE</MAINENTRYDEPLOY>
    <SUBENTRYDEPLOY>FALSE</SUBENTRYDEPLOY>
    <ID> BUMHN_S001_1907_T013_N001_JPG</ID>
  </CONTROL>
  <FRONT>
    <TITLE EMP="BOLD">INDEX</TITLE>
    <SUBTITLE>NOMS D'ESPECES</SUBTITLE>
    <COMMENT>
      <P>CITEES DANS LE COURS DE L'OUVRAGE</P>
    </COMMENT>
  </FRONT>
```

Chaque entrée d'index contient 3 valeurs codifiées comme suit :

- 1) entrée d'index (libellé),
- 2) numéro de la pagination originale de l'imprimé telle qu'elle se présente dans le document,
- 3) nom du fichier numérique correspondant à cette page

Ce fichier prendra la forme, pour chaque entrée :

```
<ENTRY TYPE="SUB">
  <TITLE>ARNAGE</TITLE>
  <PAGES>
    <PAGE REF=" BUMHN_S001_1907_T013_N001_0172.JPG"
LABEL="174"/>
```

```
<PAGE REF=" BUMHN_S001_1907_T013_N001_0179.JPG"
LABEL="182"/>
<PAGE REF=" BUMHN_S001_1907_T013_N001_0189.JPG"
LABEL="192"/>
</PAGES>
```

Les balises doivent se succéder sans être séparées par des caractères (espace, tabulation, retour à la ligne, retour chariot). Les fichiers doivent être saisis sans retour à la ligne ni retour chariot.

Seuls les chiffres arabes sont acceptés dans la numérotation.

Les règles de saisie précises et les fichiers modèles seront fournies ultérieurement.

Remarques :

1) Seules les tables d'index expressément demandées par la bibliothèque font l'objet de ce traitement.

A titre indicatif, dans les Bulletins du Museum, on compte moins de 300 pages d'index d'espèces (annexe 1).

2) Il ne sera pas demandé au prestataire de créer des tables d'index absentes des documents.

5.3 Nommage du répertoire et des fichiers des tables de matières et/ou d'index (voir aussi 1.3).

1. Nom du répertoire de la table des matières et/ou des index

L'ensemble des fichiers des tables des matières et/ou des index d'un même titre (de publication en série ou de monographie) sera rangé dans un répertoire dont le nom reprend l'identifiant affecté au titre, suivi de l'extension TDM : XXXXX_TDM

Le répertoire TDM est lui-même intégré dans le répertoire racine identifiant l'ouvrage ou le fascicule correspondant (voir 1.3.3.)

Exemple : BMAZO_TDM

⇔ Répertoire contenant toutes les tables des matières et éventuellement les index du *Bulletin*, section A, Zoologie.

2. Nom des fichiers des tables des matières et/ou des index :

Les fichiers des tables des matières ont le même nom que les fascicules correspondants, suivis de l'extension correspondante : .xls ou .xml

Afin de différencier le fichier de la table des matières de celui de l'index, l'information sera ajoutée en complément du nom du fichier.

Exemple : BMAZO_S003_1972_T011_N201_tdm.xls

⇔ Nom de *la table des matières* de l'unité documentaire des *Bulletins*, section A, Zoologie, 3^e série, 1972, Tome 11, numéro 201, saisie en xls. (voir 1.3)

BMAZO_S003_1972_T011_N201_ind.xml

⇔ Nom de *l'index* de l'unité documentaire des *Bulletins*, section A, Zoologie, 3^e série, 1972, Tome 11, numéro 201, saisi en xml. (voir 1.3)

Article 6 - Prestations de transport, de conservation et de stockage

1. - Pour le transport :

Dans le cadre de l'exécution des prestations du présent marché, le titulaire s'engage à prendre en charge le transport des documents confiés pour la numérisation depuis la Bibliothèque jusqu'au lieu d'exécution des prestations, aller et retour. Le prestataire garantira les meilleurs délais et conditions de transport.

Le titulaire devra effectuer ses tournées de façon à optimiser au mieux les déplacements et limiter leurs impacts sur l'environnement.

2. - Pour les conditions de stockage et de conservation sur site :

- L'Hygrométrie, la température et la luminosité seront conformes aux règles minimum de conservation applicables aux documents fournis. Les dispositions devront être prises contre les insectes, les incendies, l'inondation et le vol.
- Surface de stockage suffisante pour entreposer les documents pendant le temps nécessaire aux traitements, à la livraison des fichiers et éventuellement jusqu'à leur retour (dans le cas où il serait expressément demandé par la Bibliothèque)

Article 7 Supports de livraison

Le prestataire fournira les disques nécessaires à l'exécution des prestations

Le support retenu est de type disque dur amovible USB2. Le titulaire prendra au moment de la livraison toutes les précautions utiles pour la protection des disques et de leur emballage de manière à éviter tout risque de détérioration.

Les disques sont accompagnés d'un bordereau de remise sur papier qui permettra l'identification et le volume de la production livrée. Le bordereau contiendra les indications suivantes :

- identifiant des supports ;
- identifiant des documents ;

Structure du disque

- Un fichier placé à la racine de l'arborescence du disque permet son identification et liste son contenu.
- La structure du disque est conforme à l'architecture signalée en annexe (arborescence des répertoires et des fichiers)
- Le disque sera étiqueté avec les contenus indiqués

Lors de la livraison, un contrôle sera effectué afin de vérifier la conformité du travail fait avec les demandes comprises dans le présent marché.

Article 8 Modalités de contrôle par le MNHN

8.1 Phase test

Pour chaque ensemble de documents, des processus de préparation, de production et de contrôle seront mis en place et vérifiés par un jeu de test préalable.

La durée de cette phase test ne pourra être supérieure à 15 jours francs.

Ce jeu de tests correspondra à un échantillon de documents défini en début de marché. Ainsi le titulaire pourra caler sa chaîne de production et le Muséum pourra vérifier la qualité de la prestation effectuée.

Au cours de cette phase, un échange entre le titulaire et le Muséum s'instaurera afin d'optimiser ce calage.

Cette phase précèdera l'exécution du marché proprement dite.

8.2 Contrôle qualité : procédure

3. Obligations du titulaire :

Il est du devoir titulaire de mettre en place un processus de contrôle de sa production qui lui permette de satisfaire aux exigences édictées dans le présent cahier des charges.

Ce contrôle concerne :

- la qualité de la saisie et de la structuration ;
- la validité des liens ;
- la qualité d'écriture sur disque
- la qualité des images

4. Contrôles effectués par la Bibliothèque du Muséum :

A réception de la fourniture, la Bibliothèque du Muséum procèdera à divers contrôles lui permettant d'apprécier la qualité de la prestation du titulaire. Deux types de contrôles seront effectués :

- Contrôle à la livraison

- Contrôle de production

5. Contrôle à la livraison :

La première opération concerne le contrôle de conformité du disque (format et structure d'enregistrement des données). Toute erreur constatée entraîne le rejet du disque.

La réfection des prestations sera à la charge du titulaire. Le délai de correction de la production erronée sera fixé d'un commun accord entre le titulaire et la Bibliothèque.

6. Contrôle de la production :

Pour les imprimés, en application des normes NF X 06-021 et NF X 06-022 (contrôle statistiques des lots), il sera pratiqué un certain nombre de contrôles visuels à partir des disques fournis par le titulaire à l'issue de sa production.

a) Erreurs de numérisation

D'une manière générale, les contrôles portent sur :

- la qualité intrinsèque des images et leur présentation,
- la dérive entre le document d'origine et sa représentation numérisée,
- la complétude et l'organisation du lot numérisé,
- le format et la structure des données.

Exemples d'erreurs majeures entraînant la non admission de la livraison :

- disque illisible
- format non respecté
- fichiers incohérents
- incohérence entre les images et les références
- page(s) ou image(s) manquante(s)
- troncature de l'information
- erreur de saisie table de correspondance

Exemples d'erreurs mineures :

- lisibilité du texte ou rendu des photographies insuffisant dû à un manque d'optimisation des réglages du numériseur,
- travers supérieurs à + ou - 1°30'
- altération des données de références fournies et mauvais renseignement des champs modifiables dans les fichiers.

Le résultat du contrôle sera notifié au prestataire dans un délai maximum de 3 semaines, à compter de la date de réception de la fourniture. Les documents invalidés au contrôle qualité seront remis à la disposition du titulaire pour réfection du travail. La réfection du travail sera à la charge du prestataire. L'ensemble considéré comme à refaire peut être le disque dur complet suivant le nombre de documents en défaut. Le délai de correction de la production erronée sera fixé d'un commun accord avec le

prestataire. La reprise devra néanmoins être effectuée dans un délai au maximum égal au délai de production initial. Cette reprise devra être faite en parallèle de la production courante et sans affecter celle-ci.

b) Qualité de la reconnaissance optique de caractères

Dans le cas de l'option OCR sans correction, si la bibliothèque choisit cette option et que le taux d'erreur constaté en cours de production est supérieur de façon récurrente à 1%, la bibliothèque se réserve le droit de renoncer à cette option.

c) Erreurs dans les TDM

Conformité au modèle :

Les fichiers saisis doivent être strictement conformes au modèle du fichier Excel fourni.

Saisie des caractères

Le taux d'erreur admissibles est fixé à 1/2000. C'est une moyenne établie sur les sous ensembles de production. Une erreur correspond à tout caractère manquant ou erroné. Un caractère est considéré comme erroné s'il a été remplacé par un autre ou affecté d'une faute de diacritique.

Saisie des informations :

La saisie des informations (titres d'articles, nom, prénom des auteurs, etc.) doit être la plus rigoureuse possible. Le prestataire doit tenir compte des remarques et observations de la Bibliothèque, pour les opérations suivantes.

Codage des liens

Aucune erreur n'est tolérée pour le codage des liens. Une erreur correspond à une incohérence entre le numéro de la page indiqué dans la table des matières et le numéro du fichier image correspondant.

Tout lien oublié est considéré comme une erreur.