# Cellule MathDoc

# Projet NUMDAM

## Numérisation de Documents Anciens Mathématiques

CAHIER DES CLAUSES TECHNIQUES PARTICULIERES

*Thierry Bouche, chef de projet*

**Cellule MathDoc, UMS 5638 CNRS UJF**
**Université Joseph Fourier**

Pierre Bérard, Directeur
Laurent Guillopé, Directeur-adjoint

Adresse géographique : Bâtiment CICG
351 avenue de la Bibliothèque
Domaine universitaire F-38402 Saint Martin d'Hères

Adresse postale : Bâtiment CICG
BP 53 F-38041 Grenoble Cedex 9

Tél : + 33 (0)4 76 63 56 36 / Fax : + 33 (0)4 76 63 56 11

**ums5638@mathdoc.ujf-grenoble.fr**
**http://www-mathoc.ujf-grenoble.fr/**

# SOMMAIRE

# 1. CADRE DE L'OPÉRATION

## 1.1. Présentation générale

La Cellule MathDoc (UMS 5638 – Université Joseph Fourier & CNRS) a été chargée de piloter une opération de numérisation des revues de mathématiques publiées en France jusqu'à la fin du XX$^{e}$ siècle.

La présente opération, qui constitue la première vague du projet global, doit être réalisée d'ici fin 2003 ; elle porte sur les cinq revues principales des origines jusqu'à l'an 2000 compris, soit un total d'environ 205 000 pages (140 000 pour la tranche ferme, 65 000 pour la tranche conditionnelle, voir § 1.2 et Annexe 1). D'autres campagnes suivront dans le but d'atteindre à terme l'exhaustivité.

Les objectifs de cette campagne de numérisation sont d'une part l'archivage sur support électronique des volumes existants, et d'autre part la mise à disposition de ces données au profit de la communauté scientifique.

Dans cette optique, le travail à accomplir sera décomposé en trois temps :

1. Collation et recensement de collections complètes pour chacune des revues concernées.

2. Numérisation de haute qualité des originaux, création et alimentation d'une base de données permettant l'indexation des articles, reconnaissance optique du texte des articles permettant la recherche plein-texte de leur contenu, plus en option une indexation des références citées.

3. Mise en place d'une plate-forme logicielle exploitant conjointement les images et la base de données, permettant une consultation sur la toile.

La Cellule MathDoc prend en charge les étapes 1 et 3, en collaboration avec les éditeurs des revues. La deuxième fait l'objet du présent appel d'offres.

En bout de chaîne, l'utilisateur final devra disposer d'une interface conviviale pour accéder aux articles, d'un affichage ergonomique sur écran, et de la capacité d'imprimer à haute définition les pages choisies.

## 1.2. Composition du futur marché

Le marché comprendra :

- Une tranche ferme portant sur le traitement successif de quatre revues : *Annales de l'institut Fourier* (AIF), *Publications mathématiques de l'institut des Hautes Études scientifiques* (PMIHES), *Bulletin de la Société mathématique de France* (BSMF) et son supplément, les *Mémoires de la Société mathématique de France* (MSMF), et *Journal des équations aux dérivées partielles* (JEDP).

- Une tranche conditionnelle qui porte sur le traitement d'une autre revue scientifique (RS).

Les sociétés doivent répondre à l'ensemble des lots. La réalisation des lots sera successive.

### 1.3. Le cadre organisationnel de l'opération de numérisation

La Cellule MathDoc prend en charge un certain nombre de travaux :

- Travaux préparatoires :
  - la collation des revues pour en vérifier l'exhaustivité, l'analyse a porté sur la totalité des fascicules ;
  - la création d'une base de données préliminaire recensant les articles de chaque revue ;
  - des comptages et des statistiques sur le contenu des fascicules ; sauf pour les AIF, il s'agit d'une analyse sur un échantillon des fascicules à traiter ;
- Admission des travaux : à l'issue des opérations de conversion, la Cellule MathDoc effectuera des contrôles pour vérification de la prestation, réalisés selon des modalités détaillées au chapitre 6.

## 2. DESCRIPTION DES DOCUMENTS À TRAITER

Toutes les revues comportent des formules mathématiques, des tableaux, des schémas, des dessins au trait et des figures en niveau de gris, plus rarement (quelques unités) : des hors-textes, photos (parfois en couleur) ou planches.

Leurs principales caractéristiques sont les suivantes :

| Revue | AIF | BSMF | MSMF | PMIHES | JEDP | RS |
|---|---|---|---|---|---|---|
| **Nom(s) et ISSN(s)** | Annales de l'Institut Fourier<br>[0373-0956]<br>Suite partielle de :<br>Annales de l'Université de Grenoble. Section sciences mathématiques et physiques<br>[0765-8834] | Bulletin de la Société mathématique de France<br>[0037-9484] | Mémoires de la Société mathématique de France<br>[0249-633X] | Publications mathématiques de l'IHES<br>[0073-8301] | Actes des Journées équations aux dérivées partielles,<br>Saint-Jean de Monts<br>(titre variable d'un tome à l'autre) | Revue scientifique supplémentaire |
| **Depuis** | 1949 | 1872 | 1964 | 1959 | 1975 | 1864 |
| **Période de numérisation** | 1949-2000 | 1872-2000 | 1964-2000 | 1959-2000 | 1975-2000 | 1864-1998 |
| **Nombres de pages** | 51 000 | 45 500 | 17 500 | 16 500 | 5 500 | ~ 67 000 |
| **Nombres d'articles** | 1 774 | 2 500 | 394 | 334 | 470 | ~ 1 750 |

Les revues sont présentées en détail dans l'annexe 1.

N.B. : Compte tenu des différences très marquées d'un article à l'autre, il est extrêmement difficile de donner des fréquences fiables (ou des moyennes pertinentes) sur le nombre d'équations par pages, l'utilisation de figures au trait ou en niveau de gris. On donne donc dans l'annexe 1, soit un chiffre exact portant sur la totalité de la revue, soit une fourchette ou une moyenne estimée à partir d'un échantillon jugé représentatif « à l'œil nu ». Le nombre moyen de caractères par page peut être estimé à 85 % de la valeur haute, qui a été calculée sur des pages très pleines (pas de figures, de titre ou d'équations centrées).

Un jeu d'échantillons numériques sera mis à la disposition des prestataires, qui pourront, le cas échéant, accéder aux collections destinées à la numérisation à Grenoble, dans les locaux de la Cellule MathDoc, sur rendez-vous.

## 3. PRESTATIONS DEMANDÉES

La prestation porte sur :

- la numérisation intégrale des fascicules ;
- la reconnaissance optique des caractères d'une partie du texte numérisé ;
- l'alimentation d'une base de données ;
- la fourniture des données sur CD-R ;
- la restitution des revues reliassées.

Pour chacune des revues, la prestation doit aboutir à la livraison de trois fournitures distinctes :

- deux séries de fichiers en mode image (TIFF) destinés à la conservation, correspondant au fac similé intégral des revues :
  - l'une segmentée en unités physiques, par pages ;
  - l'autre en unités logiques (articles ou autre élément logique tel que pages de couverture…) ;
- une série de fichiers structurés en mode texte :
  - un fichier XML par volume physique comportant les métadonnées associées à chaque article ;
  - un fichier XML par article comportant le texte reconnu (plein-texte) ;
- une ou plusieurs série de fichiers (multipages) destinés à l'utilisateur final :
  - un fichier PDF par article comportant le texte reconnu de façon invisible, uniquement pour permettre les recherches ;
  - tout autre format de fichier image maîtrisé par l'opérateur, offrant une ergonomie, un poids et une qualité de restitution compétitive par rapport au PDF (DjVu, …). Il est demandé au prestataire d'argumenter ses propositions, de fournir des exemples, et d'évaluer le surcoût engendré par la fourniture de fichiers sous différents formats PDF, etc.

### 3.1. Numérisation et fourniture des fichiers image

Une collection complète de chaque revue sera adressée à l'opérateur. Les collections fournies pourront être massicotées, à l'exception d'un volume des PMIHES; il est demandé au prestataire de les restituer reliassées. Certains tomes ou fascicules des *Publications mathématiques de l'IHES* étant des exemplaires uniques, ils seront fournis déliassés ; un cahier (24 p.) de l'un d'entre eux ayant totalement disparu, il sera numérisé à partir d'un exemplaire relié à retourner en l'état, sans massicotage, mais dont l'ouverture à 180° est possible (tome relié sous couverture cartonnée).

Cette collection sera accompagnée d'une base de données préliminaire réalisée par la Cellule MathDoc attribuant en particulier à chaque volume physique et chaque unité logique un identifiant unique auquel il reviendra à l'opérateur d'associer les données numériques en retour.

Les documents livrés (fascicules, volumes reliés, volumes isolés) seront numérisés de la première à la dernière page physique, permettant ainsi leur reconstitution physique totale si nécessaire. Les couvertures seront également numérisées, en couleur, y compris les pages intérieures qui portent parfois l'ours du tome. Cependant, les couvertures des tomes 32 et suivants des PMIHES, qui sont des reliures rigides, ne seront pas numérisées. L'opérateur attribuera à chaque page physique numérisée un numéro identifiant qu'il pourra définir selon ses propres règles, qui seront clairement explicitées.

L'ensemble des pages numérisées d'un volume physique sera segmenté selon les unités logiques (articles, communications) identifiées dans la base de données.

Dans certaines revues (systématiquement dans le BSMF avant 1926, 53 tomes sur 128, segments 1 et 2, cf. annexe 1 ; exceptionnellement dans le cas d'annexes ou d'addendums rédigés comme un article en soi, inséré immédiatement à la suite de l'article auquel il se réfère dans les autres revues), les articles et communications peuvent s'enchaîner sur une même page. Il est donc demandé au prestataire d'éliminer des unités logiques les éléments (début ou fin d'un autre article ou publicité…) qui ne relèvent pas de ladite unité logique. Le bulletin de la SMF comporte également une section « vie de la société » qui contient des communications courtes. Cette section, qui représente une proportion d'environ 17 % des pages est dispersée (en général en quatre sections) dans les tomes 1 à 38, puis regroupée en fin de volume (avec foliotage réinitialisé à 1) entre les tomes 39 à 78. Nous prévoyons de traiter ces parties comme une seule unité logique du point de vue de la segmentation des images (un unique fichier multipage regroupera la totalité de chaque section de ce type), mais la base de données préparée par MathDoc répertoriera chaque communication individuellement en tant qu'article de type communication (voir annexes).

L'opérateur livrera donc d'une part une série linéaire d'images correspondant à chaque page traitée, et d'autre part une série de fichiers multipages ne comportant qu'une unité logique et nettoyés de toute information parasite si nécessaire, notamment dans le cas d'articles disposés à la suite en continu.

Le format pour les fichiers images, sauf meilleure proposition de l'opérateur, sera : TIFF monopage noir et blanc (1 bit) compressé (CCITT groupe 4) à 600 dpi optiques (non issus d'un calcul par interpolation) pour les zones de texte. Les images (schémas, photographies) en tons continus ou en couleur seront numérisées en 256 niveaux de gris ou 36 bits et compressées également.

Les spécifications sont les mêmes pour les fichiers multipages ; un format hybride autorisant de ne coder par exemple la couleur que sur la partie de l'article comportant une image en couleur sera apprécié.

Le format visuel de l'image respectera les dimensions du papier de l'original.

Les fichiers images multipages servant de base à la suite du processus, il est souhaitable qu'elles soient aussi nettes que possible, et débarrassées de tout accident dû à la numérisation ou à la qualité des originaux. Il est donc demandé à l'opérateur de mettre en œuvre des fonctions de redressement d'image, de nettoyage des bords et des traces parasites.

### 3.2. Fourniture des fichiers en mode caractère

Les recherches menées par les utilisateurs porteront sur les notices de la base de données dont un élément (conservé dans un fichier propre plein-texte XML, voir l'annexe 3) sera le texte reconnu des articles, ci-après dénommé : plein-texte, structuré de telle sorte que la page sur laquelle un mot a été reconnu soit connue. On demande par ailleurs que le plein-texte soit présent dans le fichier utilisateur de telle sorte qu'il soit possible de chercher un mot dans ces fichiers. Les formats PDF ou DjVu offrent cette possibilité. Les fichiers à ces formats devront être protégés contre la copie du plein-texte (texte caché interdit en copie, similaire à ce que l'on peut produire avec Acrobat Capture).

Il est donc demandé au prestataire de fournir un fichier texte des articles et autres unités logiques assimilées (cf. annexe 3) ; les communications contenues dans la section « Vie de la société » du BSMF ne feront pas l'objet de ce traitement, non plus que les sommaires ou toute partie de la revue qui n'est pas considérée comme un article au sens de l'annexe 3 (le taux de pages non soumises à ce traitement est de l'ordre de 10 % pour les AIF, 5 % pour les autres, à quoi il faut ajouter environ 5 000 pages pour les sections « Vie de la société » du BSMF des tomes 1 à 78). Les parties à traiter sont exclusivement le texte lui-même, à l'exclusion des formules mathématiques, des tableaux, des schémas, des dessins au trait et des figures.

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 si celle-ci offre une meilleure qualité pour un coût identique. Le soumissionnaire indiquera la méthode retenue, les moyens mis en œuvre pour ce traitement ainsi que le niveau de qualité (nombre d'erreurs moyen par mille de caractères) sur lequel il s'engage.

### 3.3. Alimentation de la base de données

La Cellule MathDoc fournira une version préliminaire de la base de données des unités logiques au format XML décrit dans l'annexe 3. Il est demandé au prestataire de vérifier la validité des informations fournies (ce sont les champs munis d'un astérisque dans l'annexe 3) par confrontation avec les originaux, et de remplir les champs restants (les champs non marqués d'un F sont obligatoires, les autres dépendent de la nature du contenu). La base comprendra les métadonnées associées à chaque volume physique, sous forme de notices XML selon une DTD fournie par la Cellule MathDoc, cf. Annexe 3.

L'opérateur décidera d'un système de nommage et d'un système de fichiers pour stocker de façon cohérente et prédictible les fichiers images (mono et multipages). Les unités logiques seront repérées par un identifiant fourni par la Cellule MathDoc, la base de données des unités logiques fournira la correspondance entre identifiants logiques et fichiers physiques.

### 3.4. Traitement des bibliographies (Option)

Il est demandé au prestataire de reprendre les références citées en bibliographie des articles selon la DTD fournie (annexe 3). Le devis devra proposer deux niveaux de finesse :

- **Option A :** Le champ **article** comprendra un champ **bibliographie** dans lequel chaque entrée sera identifiée (le texte complet de chaque référence bibliographique sera inséré dans un champ **bibitem**).
- **Option B :** Le champ **article** comprendra un champ **bibliographie** dans lequel chaque entrée sera identifiée comme dans le cas de l'option A. À l'intérieur de chaque champ **bibitem**, les noms d'auteurs, le titre de l'article cité, et sa date de publication, seront identifiés par une balise spécifique. Les éléments restants ne seront pas effacés.

La qualité des informations doit être comparable à la saisie en vue d'une nouvelle édition. Le soumissionnaire indiquera la méthode retenue, les moyens mis en œuvre pour ce traitement ainsi que le niveau de qualité (nombre d'erreurs moyen par mille de caractères) sur lequel il s'engage. Les bibliographies à prendre en compte sont exclusivement celles situées en fin d'article : les références bibliographiques citées dans les textes ou en note de bas de page ne sont pas à traiter. Les bibliographies comportent en moyenne une vingtaine de références ; environ 4 700 articles sur 7 250 (dont 3 800/5 500 pour la tranche ferme) comportent une bibliographie.

# 4. ORGANISATION DE LA PRESTATION

## 4.1. Organisation des traitements

Ils devront comprendre :

- l'élaboration de consignes détaillées de saisie et de traitement ;
- la réalisation d'un banc d'essai. Au démarrage de l'opération, un banc d'essai sur un pourcentage de pages correspondant à l'importance de la revue devra permettre de valider les consignes de saisie et d'élaborer des consignes complémentaires de traitement des anomalies détectées en cours de saisie. Le nombre de pages à traiter sera de 800 pages par revue (AIF, PMIHES, BSMF), sauf pour JEDP et MSMF : 300 pages. Le banc d'essai permettra également de valider le format de saisie et le chargement dans le système informatique de la Cellule MathDoc ;
- le traitement des revues en appliquant les consignes propres à chacune des revues.

Afin de faciliter l'exécution des prestations la Cellule MathDoc s'engage à fournir toutes les indications nécessaires à l'affinement des consignes ou à la résolution de cas spécifiques (appelés par la suite dans ce texte « anomalies »). À l'issue de la saisie de chaque ensemble d'environ 5 000 pages, un listing d'anomalies sera édité par le prestataire de saisie et envoyé pour résolution ; les réponses seront adressées au prestataire dans un délai d'une semaine pour les cas simples. Les cas complexes (indication de correction après consultation de l'éditeur…) seront regroupés pour être traités à la fin des opérations de saisie.

## 4.2. Suivi de la réalisation

Le chef de projet responsable du suivi général de l'opération sera le correspondant du titulaire pour toutes les questions propres aux traitements.

Une revue d'avancement sera faite tous les mois par le responsable de projet du titulaire et le chef de projet de la Cellule MathDoc. Des réunions intermédiaires pourront être demandées par la Cellule MathDoc si besoin est.

Le chef de projet de la Cellule MathDoc pourra être assisté pour toutes les vérifications et contrôles de qualité par le prestataire de son choix.

### 4.3. Calendrier

Dans les jours qui suivent la notification du marché, les parties concernées décident en concertation d'une date de démarrage de l'opération de numérisation. À partir de cette date, le calendrier prévisionnel général est le suivant :

| *Date à partir du démarrage* | *Tâche* | *Remarques* |
|---|---|---|
| D | Élaboration des spécifications du banc d'essai | délai : 1 mois |
| D + 1 mois | Fourniture des spécifications, démarrage du banc d'essai | délai : 1 mois |
| D + 2 mois | Fourniture du banc d'essai | contrôle par la Cellule MathDoc des spécifications et du banc d'essai |
| D + 2,5 mois | Validation du banc d'essai | |
| D + 3 mois | Lancement du traitement des AIF | délai : 1,5 mois ; livraison en deux lots d'environ 25 000 pages |
| D + 4,5 mois | Fin du traitement des AIF | |
| D + 5 mois | Lancement du traitement des PMIHES | délai : 3 semaines ; livraison en un seul lot |
| D + 6 mois | Fin du traitement des PMIHES | |
| D + 6 mois | Lancement du traitement du JEDP | délai : 3 semaines ; livraison en un seul lot |
| D + 7 mois | Fin du traitement du JEDP | |
| D + 7,5 mois | Lancement du BSMF | délai : 2 mois ; livraison en deux lots |
| D + 9,5 mois | Fin du traitement du BSMF | |
| D + 10 mois | Lancement des MSMF | délai : 1 mois ; livraison en un seul lot |
| D + 11 mois | Fin du traitement des MSMF | |

Il est demandé aux sociétés de proposer un calendrier détaillé pour le traitement de chacune des revues et un calendrier général pour l'ensemble du projet. Elles peuvent proposer un calendrier différent de celui présenté ci-dessus, sous réserve de prendre en compte les contraintes propres aux différents traitements et de justifier les propositions de modification.

La tranche conditionnelle (RS) suivra un calendrier similaire à celui du BSMF à l'issue de la tranche ferme. La livraison se fera en 3 lots sur 2 mois. La notification de la tranche conditionnelle interviendra avant le traitement des MSMF (D+10 mois selon le calendrier ci-dessus).

À chaque livraison de lot, la Cellule MathDoc lance les contrôles dès réception et les achève dans un délai de trois semaines. La recette, dont le montant est proportionnel au volume traité, est liquidée à l'issue des contrôles jugés satisfaisants.

### 4.4. Les résultats attendus de la prestation

Les sociétés devront fournir :

- le dossier des consignes détaillées de traitement (y compris sous forme numérique). Ces consignes feront partie du banc d'essai, qui permettra de les valider ;

Chaque lot livré sera accompagné

- du dossier des consignes détaillées de traitement particulières pour ce lot, s'il y a lieu ;
- du listage (y compris sous forme numérique) de l'ensemble des données traitées. Ces listes comporteront l'identifiant et les noms des fichiers informatiques correspondants ;
- des fichiers demandés sur CD-R multisession (en double exemplaire) accompagnés des consignes de chargement. Il sera préférable de livrer les fichiers d'archivage (TIFF) et les fichiers d'exploitation (XML, PDF…) sur deux supports différents ;
  - par page numérisée, y compris les couvertures : un fichier TIFF,
  - par volume physique : un fichier XML comprenant les métadonnées correspondantes selon la DTD fournie par la Cellule MathDoc,
  - par article :
    - un fichier TIFF multipage,
    - un fichier XML contenant le texte reconnu,
    - un fichier PDF contenant l'image des pages et le texte reconnu « caché » sur la page correspondante,
    - un ou plusieurs fichiers compétitifs par rapport au fichier PDF, en terme de poids, d'ergonomie ou de fonctionnalités ;
- des originaux papier reliassés.

Ces fournitures doivent contenir les informations nécessaires au contrôle d'exhaustivité et de qualité.

# 5. CONDITIONS DE RÉALISATION DE LA PRESTATION

## 5.1. Engagements de la Cellule MathDoc

La Cellule MathDoc s'engage à fournir au titulaire du marché pour chaque lot tous les éléments nécessaires à la réalisation de la prestation, à savoir :

- l'ensemble des informations nécessaires à la bonne réalisation des traitements
  - validation des spécifications,
  - réponses aux questions d'ordre scientifique ou technique liées aux fichiers,
- la désignation d'un chef de projet chargé de suivre l'opération,
- la préparation et la mise à disposition des revues, conformément au calendrier prévu et accepté par le prestataire,
- les contrôles et l'admission des lots (conformément au chapitre 6 du présent CCTP).

## 5.2. Engagements du futur titulaire du marché

### *5.2.1. Obligation de résultats*

Le titulaire sera soumis à une obligation de résultats. Il est responsable de la qualité des traitements et doit livrer un produit conforme aux critères techniques et de qualité définis dans le CCTP, et par ses propositions lorsqu'elles ont été retenues, pour chaque type de fichier.

### *5.2.2. Autres engagements*

Respect du calendrier annoncé.

La société s'engagera dans sa réponse sur un délai maximal pour le traitement des revues. En l'absence de retard imputable au maître d'ouvrage (non respect des délais de contrôle et de traitement des anomalies), ce délai doit être respecté sous peine de se voir appliquer les pénalités de retard prévues au CCAP.

Les tarifs proposés pour la prestation seront valables 120 jours à compter de la date limite de dépôt des plis en réponse au présent appel d'offre.

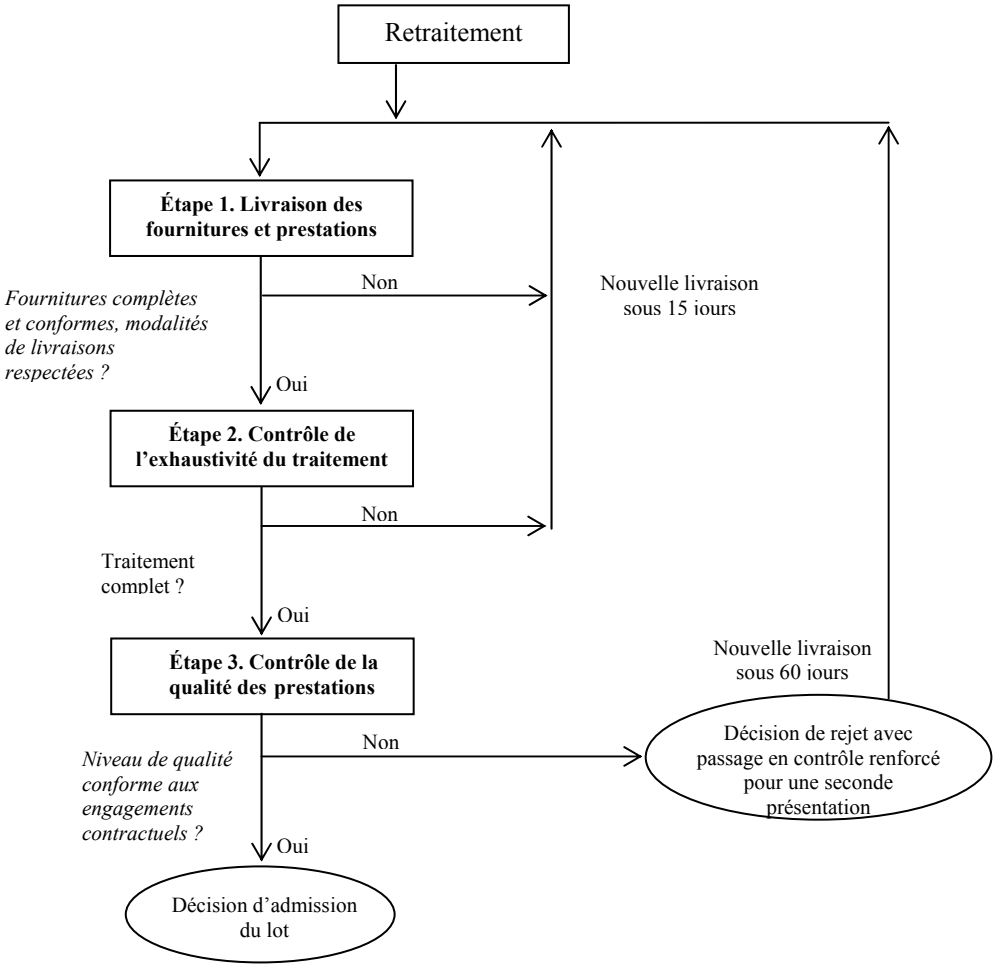
## 6. VÉRIFICATIONS, VALIDATIONS ET ADMISSION

### 6.1. Vérifications

Les prestations seront soumises à des vérifications destinées à constater qu'elles répondent aux spécifications du cahier des charges.

Les contrôles seront réalisés selon le schéma de traitement défini ci-dessous. Les validations se feront par sous-lot livré et déclencheront la facturation et le paiement des prestations liées à ce sous-lot.

*Présentation de la procédure de contrôle*




#### *6.1.1. Problématique du contrôle*

Définition de la non-conformité :

Résultat différent de celui que laisse prévoir l'application des consignes de saisie, résultat non justifié par les comptes rendus de saisie faits par le prestataire.

Dans le cas du contrôle d'exhaustivité, la non-conformité s'exprime en termes de différence de nombre de pages ou nombre d'unités logiques. Le nombre de fichiers informatiques est différent de celui calculé par application des consignes de saisie, sans que les comptes rendus des opérations de traitement justifient cette différence.

Dans le cas du contrôle de qualité, la non-conformité s'exprime en termes de différence entre le fichier informatique et le résultat attendu par application des consignes de traitement. La différence peut porter sur :

- le zonage (la zone de saisie d'une information n'est pas celle prévue par les consignes), sur le contenu d'une zone (l'information a été mal retranscrite) ;
- la lisibilité : mauvaise orientation ou redressement de l'image, image partiellement numérisée…
- le nombre de caractères erronés sur les zones reprises en OCR.

### *6.1.2. Niveaux d'exhaustivité et de qualité attendus*

Tout CD-R qui ne pourra être chargé sera immédiatement rejeté par la Cellule MathDoc.

Le niveau d'exhaustivité demandé est de 100%, c'est-à-dire que toutes les pages et unités logiques doivent être traitées.

Les traitements feront l'objet d'un contrôle de qualité selon une approche statistique. Le soumissionnaire exprimera précisément dans sa proposition les taux de qualité sur lesquels il s'engage quant aux différents traitements : OCR corrigé et non corrigé, segmentation, alimentation de la base de données selon les DTD définies. Il peut présenter différentes variantes en termes de niveau de qualité et de prix.

Les contrôles seront effectués selon les niveaux de qualité retenus.

### *6.1.3. Les étapes du contrôle*

Les différentes étapes du contrôle suivront la procédure décrite ci-dessous.

Pour chaque type de contrôle, en cas de non-conformité, une nouvelle livraison devra être effectuée dans un délai de quinze jours par le prestataire.

**- Vérifications quantitatives**

Le contrôle ***d'exhaustivité*** sera effectué par comparaison entre le nombre de pages et d'unités logiques fournies par le prestataire d'une part et le nombre de pages et d'unités logiques attendus d'autre part (nombre établi sur la base du recensement préparatoire fait par la Cellule MathDoc et des identifications des unités logiques dans la base de données transmise au prestataire en début de traitement).

**- Vérifications qualitatives**

Pour les contrôles de qualité effectués par échantillonnage, la qualité sera contrôlée sur les pages et unités logiques livrées par le prestataire par comparaison avec les originaux selon des règles de contrôle statistique définies principalement dans les normes AFNOR X06-021, X06-022, X06-028.

Les contrôles de qualité feront l'objet d'un premier contrôle en mode normal et en cas de rejet, d'un contrôle en mode renforcé lors de la seconde présentation.

En ce qui concerne la qualité de numérisation des pages, les modalités des contrôles et les tailles des échantillons en contrôle normal et renforcé sont les suivants :

| ***Lot (1)*** | ***Échantillon (pages) contrôle normal*** | ***Critère de rejet en contrôle normal (2)*** | ***Échantillon (pages) contrôle renforcé*** | ***Critère de rejet en contrôle renforcé (2)*** |
|---|---|---|---|---|
| *AIF* | *315* | *A = 4 R = 5* | *500* | *A = 5 R = 6* |
| *BSMF* | *315* | *A = 4 R = 5* | *500* | *A = 5 R = 6* |
| *MSMF* | *315* | *A = 4 R = 5* | *500* | *A = 5 R = 6* |
| *PMIHES* | *315* | *A = 4 R = 5* | *500* | *A = 5 R = 6* |
| *JEDP* | *200* | *A = 3 R = 4* | *315* | *A = 4 R = 5* |
| *RS* | *315* | *A = 4 R = 5* | *500* | *A = 5 R = 6* |

*(1) 2 lots pour AIF et BSMF ; 1 lot pour MSMF, PMIHES et JEDP ; 3 lots pour RS.*

*(2) A = admission , R = rejet.*

En ce qui concerne la qualité des unités logiques, associées à des fichiers multipages, les contrôles porteront sur a) la segmentation, b) les métadonnées XML, c) le plein-texte. Les modalités des contrôles et les tailles des échantillons en contrôle normal et renforcé sont les suivants :

| ***Lot (3)*** | ***Échantillon (UL) contrôle normal*** | ***Critère de rejet en contrôle normal (4)*** | ***Échantillon (UL) contrôle renforcé*** | ***Critère de rejet en contrôle renforcé (4)*** |
|---|---|---|---|---|
| *AIF* | *80* | *A = 1 R = 2* | *125* | *A = 1 R = 2* |
| *BSMF* | *125* | *A = 1 R = 2* | *200* | *A = 1 R = 2* |
| *MSMF* | *50* | *A = 1 R = 2* | *80* | *A = 1 R = 2* |
| *PMIHES* | *50* | *A = 1 R = 2* | *80* | *A = 1 R = 2* |
| *JEDP* | *50* | *A = 1 R = 2* | *80* | *A = 1 R = 2* |
| *RS* | *80* | *A = 1 R = 2* | *125* | *A = 1 R = 2* |

*(3) 2 lots pour AIF et BSMF ; 1 lot pour MSMF, PMIHES et JEDP ; 3 lots pour RS.*

*(4) A = admission , R = rejet.*

### 6.2. Validations

Les validations seront le fait de la Cellule MathDoc. Elles seront réalisées dans les 3 semaines suivant la remise des fournitures liées au traitement d'un lot et déclencheront la facturation et le paiement des prestations correspondantes.

Un lot rejeté pour des erreurs imputables au prestataire doit être traité à nouveau par le prestataire et à ses frais. Un lot accepté permet la facturation de la prestation correspondante.

En cas de rejet ou de demande de livraison de complémentaire, les délais prévus pour une nouvelle livraison à effectuer par le prestataire sont de 15 jours.

### 6.3. Admission du marché

Cette décision unique pour chaque lot sera prononcée par la personne publique responsable du marché après que l'ensemble des vérifications aura été considéré comme positif.

A l'admission, il y aura transfert de propriété du produit, c'est-à-dire des données saisies par le prestataire, au Maître d'ouvrage.