

Cellule MathDoc
Programme NUMDAM
Numérisation de Documents Anciens Mathématiques
Phase II

CAHIER DES CLAUSES TECHNIQUES PARTICULIÈRES

Thierry BOUCHE, chef de projet
Hélène BÉGNIS, ingénieur d'études

Cellule MathDoc, UMS 5638 CNRS UJF
Université Joseph Fourier

Yves Laurent, 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

SOMMAIRE	2
1. CADRE DE L'OPÉRATION	3
1.1. <i>Présentation générale.....</i>	<i>3</i>
1.2. <i>Composition du futur marché.....</i>	<i>3</i>
1.3. <i>Le cadre organisationnel de l'opération de numérisation.....</i>	<i>4</i>
2. DESCRIPTION DES DOCUMENTS À TRAITER.....	5
3. PRESTATIONS DEMANDÉES.....	6
3.1. <i>Généralités.....</i>	<i>6</i>
3.2. <i>Traitements.....</i>	<i>6</i>
3.3. <i>Numérisation et fourniture des fichiers.....</i>	<i>7</i>
3.3.1. <i>Format et qualité des fichiers images monopages.....</i>	<i>7</i>
3.3.2. <i>Traitement des articles.....</i>	<i>7</i>
4. ORGANISATION DE LA PRESTATION.....	9
4.1. <i>Organisation des traitements.....</i>	<i>9</i>
4.2. <i>Suivi de la réalisation.....</i>	<i>9</i>
4.3. <i>Déroulement des opérations.....</i>	<i>10</i>
4.4. <i>Les résultats attendus de la prestation.....</i>	<i>10</i>
5. CONDITIONS DE RÉALISATION DE LA PRESTATION	11
5.1. <i>Engagements de la Cellule MathDoc.....</i>	<i>11</i>
5.2. <i>Engagements du futur titulaire du marché.....</i>	<i>11</i>
5.2.1. <i>Obligation de résultats.....</i>	<i>11</i>
5.2.2. <i>Autres engagements.....</i>	<i>11</i>
6. VÉRIFICATIONS, VALIDATIONS ET ADMISSION	12
6.1. <i>Vérifications.....</i>	<i>12</i>
6.1.1. <i>Problématique du contrôle.....</i>	<i>13</i>
6.1.2. <i>Niveaux d'exhaustivité et de qualité attendus.....</i>	<i>13</i>
6.1.3. <i>Les étapes des contrôles.....</i>	<i>13</i>
6.1.4. <i>Nomenclature des erreurs.....</i>	<i>14</i>
6.2. <i>Validations.....</i>	<i>16</i>
6.3. <i>Admission du marché.....</i>	<i>16</i>
Cahier des charges NUMDAM phase II - Liste des annexes.....	17

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 de périodiques et de collections de mathématiques publiés en France jusqu'à la fin du XX^e siècle.

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.

La présente opération, qui constitue la deuxième vague du projet global, doit être réalisée d'ici fin 2006 ; elle porte sur plusieurs périodiques importants des origines jusqu'à l'an 2000 compris, ainsi que d'autres publications en série, comme des actes de séminaire, pour un total compris entre 50 000 et 400 000 pages.

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

1. Collation et recensement des collections de documents à numériser ;
2. Numérisation de haute qualité des originaux, création et alimentation d'une base de données permettant l'indexation du contenu des volumes, reconnaissance optique du texte des articles permettant la recherche plein-texte de leur contenu, balisage 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 documents. La deuxième fait l'objet du présent appel d'offre.

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 présent marché est un marché « à bons de commande ». Il sera constitué du traitement de plusieurs ensembles de documents d'environ 5 000 à 25 000 pages chacun, pour lesquels un même ensemble de traitements, parmi ceux décrit au chapitre 3 seront appliqués. Certains de ces traitements sont communs à tous les ensembles de documents, d'autres sont appliqués au cas par cas sur instruction de la Cellule MathDoc.

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 périodiques ;
 - Dans certains cas, la création d'une base de données préliminaire recensant le contenu des volumes.
- ❖ 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

Les documents à traiter sont des publications en série (publication qui se succède dans le temps pendant une durée qui n'est pas limitée à l'avance) composées essentiellement d'articles de recherche en mathématiques. À ce titre, toutes les séries comportent des formules de mathématiques, des tableaux, des dessins et des figures, plus rarement : des hors-texte, des photos, ...

La période de publication des documents remonte rarement avant 1800. L'essentiel des documents à traiter sont publiés au XX^e siècle.

Un échantillon des collections destinées à la numérisation sera fourni aux soumissionnaires en vue d'un test de capacité à produire en conformité aux présentes clauses techniques. Ce test doit être réalisée dans les conditions de la production.

Ce test portera sur **un volume des Annales de l'Institut Fourier**, tome 52, fascicule 2, année 2002 (environ 380 pages) et sur **un volume du Séminaire Goulaouic-Schwartz**, année 1976-1977 (environ 300 pages). Ces volumes devront être traités de la manière suivante :

- *Annales de l'Institut Fourier, tome 52, fascicule 2, année 2002 :*
 1. Scanner à livre ouvert (c'est-à-dire sans massicotage) ;
 2. Cataloguer le volume puis chaque article en suivant la DTD annexée au présent CCTP (annexe 3).
- *Séminaire Goulaouic-Schwartz, année 1976-1977 :*
 1. Scanner feuille à feuille (ce volume est massicotable) ;
 2. Compléter le catalogue XML fourni par la Cellule MathDoc (sur la disquette jointe à l'acte d'engagement) conformément à la DTD annexée au présent CCTP (annexe 3).

Les volumes utilisés pour ce test devront être restitués reliassés à la Cellule MathDoc, en même temps que les résultats des traitements au plus tard le 13 mars 2003.

À titre d'exemple, un échantillon de fichiers produits lors de la phase I du programme NUMDAM pourra être fourni sur demande.

3. PRESTATIONS DEMANDÉES

3.1. Généralités

La prestation pourra comporter les opérations suivantes :

- le massicotage des documents ;
- la numérisation de l'ensemble des pages des volumes ;
- la création ou l'alimentation d'une base de données :
 - le catalogage des volumes ;
 - la segmentation en unités logiques (articles, ...).
- le traitement des articles numérisés :
 - la création des fichiers utilisateurs multipages ;
 - la reconnaissance optique des caractères du texte de ces articles ;
 - le balisage des bibliographies de fin d'article ;
- la fourniture des données sur CD-Rom ;
- la restitution des documents, reliés, le cas échéant.

Pour chaque ensemble de documents, un bon de commande sera émis, qui précisera les traitements à effectuer. Certains de ces traitements sont systématiques, d'autres optionnels comme détaillé ci-après.

3.2. Traitements

Pour chaque ensemble de documents, la prestation doit aboutir à la livraison des trois fournitures suivantes :

- ❖ une série de fichiers en mode image (TIFF) destinés à la conservation, correspondant au fac-similé intégral de chaque page (y compris les pages blanches) des volumes traités ;
- ❖ une série de fichiers structurés en mode texte :
 - un fichier XML par volume physique comportant les métadonnées associées à chaque unité logique (article, chapitre, ...) ;
 - un fichier XML par article comportant le texte reconnu (plein-texte) ;
- ❖ deux ou trois séries de fichiers (multipages) par unité logique :
 - un fichier TIFF multipage ;
 - un fichier PDF comportant le texte reconnu de façon invisible, uniquement pour permettre les recherches ;
 - **Option** : un fichier DjVu comportant le texte reconnu de façon invisible, uniquement pour permettre les recherches.

3.3. Numérisation et fourniture des fichiers

Certains volumes fournis pourront être massicotés, d'autres devront être numérisés à livre ouvert. Les volumes massicotés devront être restitués reliés. Les autres volumes devront être restitués tels que réceptionnés.

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 2^e et 3^e de couvertures si elles comportent du texte.

L'ensemble des pages numérisées d'un volume physique sera segmenté selon les unités logiques (articles, communications).

Dans le cas où la Cellule MathDoc fournit une base de données (au format XML) cataloguant les articles, appelée dans ce document « version préliminaire de la base de données », l'opérateur l'utilisera comme référence pour la susdite segmentation. Dans le cas contraire, la version préliminaire de la base de données (au format XML) devra être créée et alimentée selon la structure de la DTD fournie en annexe 3.

Lorsque les articles ou communications s'enchaînent sur une même page, il est demandé au prestataire d'éliminer des fichiers multipages, associés à une unité logique, les éléments (début ou fin d'un article, publicité, ...) qui ne relèvent pas de ladite unité logique.

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.

3.3.1 Format et qualité des fichiers images monopages

Le format pour les fichiers images est :

- ❖ TIFF monopage noir et blanc (1 bit) compressé (CCITT groupe 4) à 600 dpi optiques (non issus d'un calcul par interpolation) pour les pages de texte, y compris pour celles contenant des graphiques.
- ❖ Les couvertures et les hors-texte en couleurs seront numérisés en couleur 24 bits à 300 dpi et compressés LZW.
- ❖ Les hors-texte en niveaux de gris seront numérisés en 256 niveaux de gris à 300 dpi et compressés LZW.

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

Les fichiers images monopages servant de base à la suite du processus, il est souhaitable qu'ils soient aussi nets que possible, et débarrassés 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.

3.3.2. Traitement des articles

Option : Création de la version préliminaire de la base de données

Pour certains ensembles de documents, il sera demandé au prestataire de produire la version préliminaire de la base de données XML qui catalogue les volumes physiques et les articles correspondants (champs munis d'un astérisque dans l'annexe 3 qui décrit la DTD « volphys »).

Alimentation de la base de données

Lorsque la Cellule MathDoc fournit la 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 renseignées (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 de données comprendra les métadonnées associées à chaque volume physique, sous forme de notices XML selon une DTD fournie par la Cellule MathDoc, cf. Annexe 3.

Dans le cas contraire, il est demandé au prestataire de cataloguer les articles en se conformant à la même DTD et de remplir tous les champs obligatoires.

Traitement des bibliographies

Il est demandé au prestataire de reprendre les références citées en bibliographie des articles selon la DTD fournie (annexe 3). Le devis devra proposer un niveau de finesse obligatoire et un deuxième niveau de finesse optionnel :

1. Le champ **article** comprendra un champ **bibliographie** dans lequel chaque référence sera identifiée par la balise **bibitem**. 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 ;
2. **Option** : l'ensemble des éléments composant une référence bibliographique sera identifié par une balise spécifique (cf. annexe 4).

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.

OCR

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 5) 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 (cf. annexe 5). 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 (texte caché 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 5).

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 estimé (nombre d'erreurs moyen par mille de caractères).

Fourniture des fichiers multipages

Pour chaque article, le prestataire fournira deux ou trois fichiers multipages :

- ❖ Un fichier TIFF multipage noir et blanc (1 bit) compressé (CCITT groupe 4) à 600 dpi optiques (non issus d'un calcul par interpolation) ;
- ❖ Un fichier PDF multipage
 - noir et blanc (1 bit) compressé (CCITT groupe 4) à 600 dpi optiques (non issus d'un calcul par interpolation) ;
 - comportant le texte reconnu de façon invisible.
- ❖ **Option** : Un fichier DjVu multipage
 - noir et blanc (1 bit) à 600 dpi optiques (non issus d'un calcul par interpolation) ;
 - comportant le texte reconnu de façon invisible, uniquement pour permettre les recherches.

4. ORGANISATION DE LA PRESTATION

4.1. Organisation des traitements

Ils devront comprendre :

- ❖ l'élaboration de consignes détaillées de saisie et de traitement ;
- ❖ l'élaboration de consignes détaillées de contrôle qualité ;
- ❖ le traitement des documents en appliquant les consignes propres à chaque ensemble de documents.

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 du traitement de chaque ensemble d'environ 5 000 pages, un listing d'anomalies sera édité par le prestataire 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.

4.2. Suivi de la réalisation

Le chef de projet désigné par la Cellule MathDoc, responsable du suivi général de l'opération NUMDAM, 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. Déroulement des opérations

Dans les jours qui suivent la notification du marché, les parties concernées rédigent en concertation les spécifications générales de la production sous un délai d'un mois ; puis conviennent d'un délai (qui ne pourra excéder 2 mois à compter de la signature des spécifications) pour la mise en place d'une chaîne de production opérationnelle.

L'opérateur précisera la localisation géographique des sites où les traitements seront effectués, en particulier la numérisation des volumes.

À chaque livraison, la Cellule MathDoc lance les contrôles dès réception et les achève dans un délai de trois semaines. La recette 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).

Chaque livraison sera accompagnée :

- ❖ du dossier des consignes détaillées de traitement particulières appliquées, 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-Rom multisession accompagnés des consignes de chargement. Les fichiers d'archivage (TIFF) et les fichiers d'exploitation (XML, PDF...) seront livrés sur deux supports différents ;
 - par page numérisée, y compris les couvertures et les pages blanches : 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,
 - **Option** : un fichier DjVu contenant l'image des pages et le texte reconnu « caché » sur la page correspondante.
- ❖ des originaux papier, reliés le cas échéant.

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 ensemble de documents 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 documents,
- ❖ les contrôles et l'admission des livraisons (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 des délais de mise en place opérationnelle de la chaîne de production.

La société s'engagera sur la durée nécessaire au traitement d'un ensemble de documents de : 5 000, 10 000 et 20 000 pages. En l'absence de retard imputable au maître d'ouvrage (non respect des délais de contrôle et de traitement des anomalies), cette durée doit être respectée sous peine de se voir appliquer les pénalités de retard prévues au CCAP.

Les tarifs proposés pour la prestation seront valables 180 jours à compter de la date limite de dépôt des plis en réponse au présent appel d'offre.

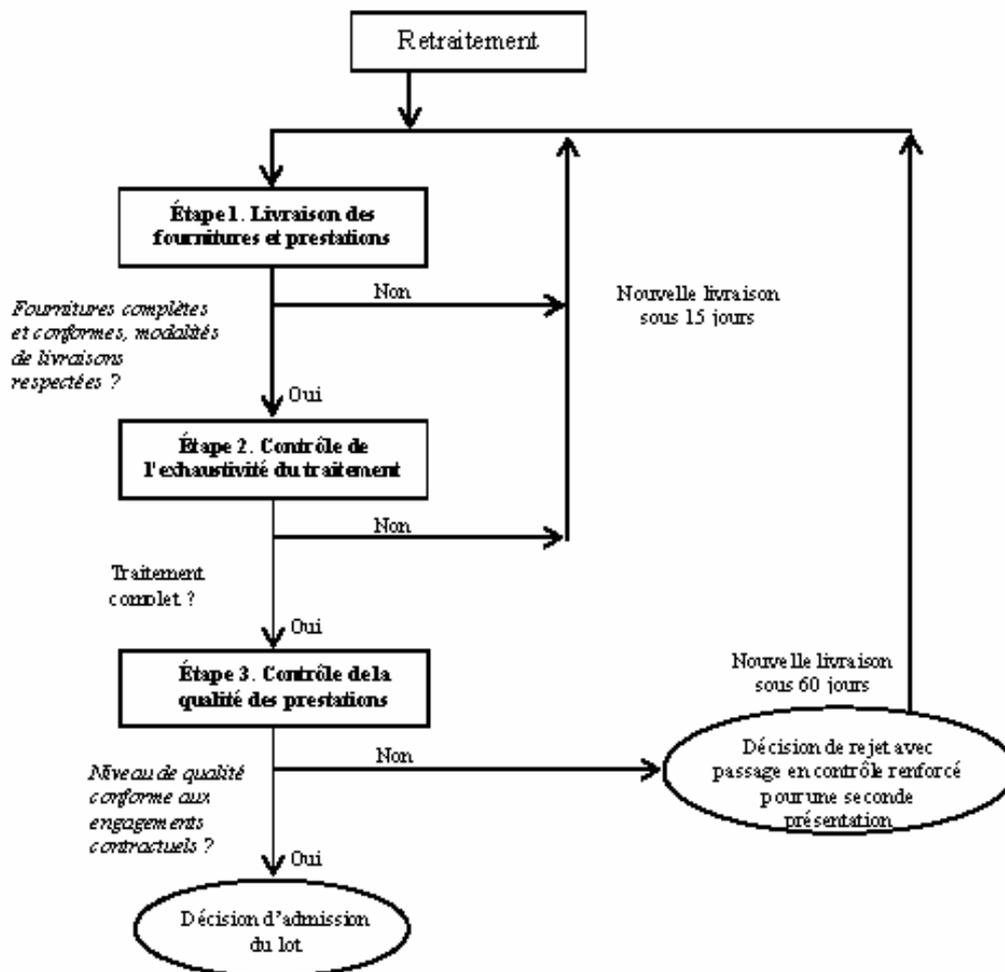
6. VÉRIFICATIONS, VALIDATIONS ET ADMISSION

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 livraison et déclencheront la facturation et le paiement des prestations liées à celle-ci.

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 traitement, résultat non justifié par les comptes rendus de traitement 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 traitement, 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 :

- ❖ l'intégrité : respect de l'information originale (segmentation des articles, pages cornées, ...)
- ❖ 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 double contrôle de qualité effectué par la Cellule MathDoc, l'un exhaustif et l'autre 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, texte et balisage des bibliographies, 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 des contrôles

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).

Vérifications qualitatives

Le contrôle exhaustif vérifie la validité de l'ensemble des fichiers livrés et la cohérence de la livraison. La non validité d'un fichier entraîne un rejet immédiat de la livraison.

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.

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.

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 :

Nombre de pages	Échantillon (pages) contrôle normal	Critère de rejet en contrôle normal (1)		Échantillon (pages) contrôle renforcé	Critère de rejet en contrôle renforcé (1)	
5 000-10 000	200	A=2	R=3	315	A=3	R=4
10 001-35 000	315	A=3	R=4	500	A=5	R=6

En ce qui concerne la qualité des unités logiques, associées à des fichiers multipages, les contrôles porteront sur a) la segmentation et l'intégrité des images 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 :

Nombre d'articles	Échantillon (articles) contrôle normal	Critère de rejet en contrôle normal (1)		Échantillon (articles) contrôle renforcé	Critère de rejet en contrôle renforcé (1)	
91-150	20	A=0	R=1	32	A=0	R=1
151-280	32	A=0	R=1	50	A=0	R=1
281-500	50	A=0	R=1	80	A=1	R=2
501-1 200	80	A=1	R=2	125	A=1	R=2
1 201-3 200	125	A=1	R=2	200	A=2	R=3

(1) A= Accepté R = rejeté

6.1.4. Nomenclature des erreurs

Les spécifications pourront apporter de plus amples précisions à cette nomenclature.

Fichiers Tiff monopages

Types d'anomalies	Décompte des erreurs par anomalie constatée
Perte d'information (caractères illisibles, page tronquée, mauvais réglage de seuil, ...)	1
Mauvaise dimension (variation supérieure à 5 % par rapport à l'original)	1
Rotation (variation supérieure à 2° par rapport à l'original)	1
Bords noirs (épaisseur supérieure à 5 % par rapport aux dimensions de l'original)	1

Fichiers Tiff multipages

Types d'anomalies	Décompte des erreurs par anomalie constatée
Page non conforme au Tiff monopage	1
Segmentation (page en trop, page en moins, page dans le désordre, page en double, page contenant du texte provenant d'un autre article)	1

Fichiers PDF et DjVu

Types d'anomalies	Décompte des erreurs par anomalie constatée
Fichier non conforme au Tiff multipage	1
Texte caché absent ou incorrect	1

Fichiers XML OCR

Types d'anomalies	Décompte des erreurs par anomalie constatée
Qualité différente du taux annoncé	1
Fichier XML non valide	1
Texte non conforme au TIFF multipage	1

Fichier XML volume

Types d'anomalies	Décompte des erreurs par anomalie constatée
Fichier XML non valide	1
Introduction d'erreur dans les données fournies	1
Non cohérence entre les métadonnées et les fichiers produits (nombre de pages, correspondance logique-physique, ...)	1
Erreur de balisage (bibliographies)	1

6.2. Validations

Les validations seront le fait de la Cellule MathDoc. Elles seront réalisées dans les 4 semaines (20 jours ouvrés) suivant la remise des fournitures constituant une livraison et déclencheront la facturation et la mise en paiement des prestations correspondantes.

Une livraison rejetée pour des erreurs imputables au prestataire doit être traitée à nouveau par le prestataire et à ses frais. Une livraison acceptée permet la facturation de la prestation correspondante.

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

6.3. Admission du marché

Cette décision unique pour chaque livraison 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.

Cahier des charges NUMDAM phase II - Liste des annexes

Annexe 1 : identifiant NUMDAM.....	18
Annexe 2 : système de fichiers	19
Annexe 3 : catalogue XML.....	20
Annexe 4 : traitement des bibliographies	25
Annexe 5 : plein-texte d'un article.....	28
Annexe 6 : cadre de réponse	29

Annexe 1 : identifiant NUMDAM

Nous utilisons actuellement un identifiant construit de la façon suivante :

- ❖ le séparateur de champs est le caractère _ (*underscore*) ;
- ❖ les champs sont, dans cet ordre :
 1. Acronyme de la revue : première lettre de chaque mot du titre (hors article défini et conjonction) ;
 2. Année (ne correspond pas forcément à l'achevé d'imprimé, dans le cas de revues ayant pris du retard ; peut prendre la forme x-y dans le cas de volumes couvrant plusieurs années) ;
 3. série ;
 4. tome (peut prendre la forme x-y dans le cas de tomes doubles, peut être précédé du préfixe S s'il s'agit d'un tome supplémentaire) ;
 5. numéro de fascicule (peut prendre la forme x-y dans le cas de fascicules doubles, peut être précédé du préfixe S s'il s'agit d'un fascicule supplémentaire) ;
 6. page initiale (il s'agit du folio, et non du numéro d'ordre linéaire à l'intérieur du volume physique considéré). Ce nombre (en chiffres arabes) sera précédé :
 - ❖ du préfixe R lorsque le folio est un chiffre romain ;
 - ❖ du préfixe S s'il s'agit d'un supplément inséré en fin d'un même volume physique, dont le foliotage est réinitialisé à 1, ou spécial (S1...) ;
 - ❖ du préfixe D dans le cas rare mais attesté d'erreurs de foliotages (doublons) qui aboutissent à la situation où plusieurs pages portent le même numéro !
 - ❖ du préfixe P s'il s'agit de pages préliminaires, avant la première page numérotée ;
 - ❖ Lorsqu'un volume n'a pas de pagination globale, les pages prenant la forme n° d'article/n° de page : on utilisera le n° d'article précédé du préfixe A ;
 - ❖ On considère que toutes les feuilles volantes [planches, hors texte, errata, etc.] sont en page zéro.
 7. numéro d'ordre sur la page.

Les champs dénués de signification sont laissés vides, mais le nombre de délimiteurs est constant. Un volume physique est identifié de la même façon, avec les champs 1-5.

Annexe 2 : système de fichiers

Système de fichiers pour fichiers monopages

L'opérateur utilisera le système de nommage et le système de fichiers suivant pour stocker les fichiers monopages :

- ❖ un dossier par revue comportant,
 - un dossier par volume physique identifié par l'identifiant NUMDAM de volume physique tel que décrit dans l'annexe 1 du présent CCTP, comportant
 - un fichier TIFF monopages par page, numéroté séquentiellement de 0001 (pour la première page intérieure) à NNNN (pour la dernière)
Exemple : page0001.tif
 page0002.tif
 - un fichier TIFF monopage par page de couverture, numéroté séquentiellement de 1 (pour la première page intérieure) à N (pour la 4^e de couverture)
Exemple : couv1.tif
 couv2.tif

Système de fichiers pour fichiers article

L'opérateur utilisera le système de nommage et de fichiers suivant pour stocker les fichiers images :

- ❖ un dossier par revue comportant,
- ❖ un dossier par volume physique identifié par l'identifiant NUMDAM de volume physique tel que décrit dans l'annexe 1 du présent CCTP, comportant
 - le catalogue XML du volume physique ;
 - un dossier par article, identifiant par l'identifiant NUMDAM de l'article, contenant,
 - un fichier TIFF multipage, identifié par l'identifiant NUMDAM de l'article
 - un fichier PDF, identifié par l'identifiant NUMDAM de l'article
 - un fichier DjVu, identifié par l'identifiant NUMDAM de l'article
 - un fichier XML, identifié par l'identifiant NUMDAM de l'article et correspondant au texte OCRisé de l'article.

Annexe 3 : catalogue XML

A. Structure des notices

Les notices seront codées en XML selon une DTD fournie par la cellule MathDoc, dont voici les spécifications. A chaque volume physique est attachée une notice décrivant l'ensemble des unités logiques qu'il contient. La notice a donc un identifiant propre, qui est celui du volume (identifiant NUMDAM de l'annexe 1 tronqué au niveau fascicule). Dans ce qui suit, les champs sont en gras, suivis de la mention **F** s'ils sont facultatifs, **R** s'ils sont répétitifs. **Les champs dotés d'un astérisque sont ceux alimentés lors de la création de la version préliminaire de la base de données (par la Cellule MathDoc ou par le prestataire)**. Le terme « version préliminaire de la base de données » dans le CCTP regroupe l'ensemble des champs décrits ci-dessous avec un astérisque. Les champs surlignés doivent être systématiquement complétés par le prestataire lors de la numérisation. Les fichiers XML sont codés en ISO-Latin-1. Les caractères n'appartenant pas à ce codage seront saisis à l'aide de références de caractères (p. ex. : ģ).

- ❖ **Notice** (description du volume physique catalogué)
 - **Identifiant NUMDAM*** (identifiant du volume)
 - **Revue***
 - **ISSN***
 - **Acronyme NUMDAM***
 - **Série* F**
 - **Tome***
 - **Numéro* F**
 - **Année***
 - **Fascicule* F**
 - **Responsable de l'édition du volume* RF** (« éditeur » des actes d'un congrès p. ex.)
 - **Éditeur*** (détenteur du copyright)
 - **Pages** (nombre total de pages physiques)
 - **Infos numérisation**
 - **Identifiant physique** (correspondance avec le répertoire correspondant dans le système de fichier « physique »)
 - **Date de création**
 - **Spécial F** (toute information concernant les caractéristiques du volume ou de sa numérisation, par exemple résolution, pages en couleur ou en niveau de gris, caractéristiques techniques, etc., ce champ est rempli de façon automatique à partir des informations fournies par le scanneur)
 - **Note F** permet de signaler les anomalies ou les particularités concernant le volume (planche, foliotage incohérent, etc.)
- ❖ **Article R**
 - **Identifiant NUMDAM*** (logique, cf. Annexe 1)
 - **Page initiale***, le contenu sera un nombre en chiffres arabes (attributs **F** : **romain, doublon, supplément, spécial**)
 - **Page finale***, le contenu sera un nombre en chiffres arabes (attributs **F** : **romain, doublon, supplément, spécial**)
 - **Numéro* F**
 - **Numéro* d'ordre sur la première page** (en général 0)
 - **Auteur* R** (Nom, Prénom)
 - **Contributeur* RF** (rôle = traducteur ou rédacteur)
 - **Traducteur* RF** (Nom, Prénom)
 - **Titre* R** (avec un attribut de **langue**, et un attribut **original** identifiant le titre original)

- **Langue*** (du corps de l'article)
- **Relations* F**
 - **Est corrigé par*** (cas d'un article suivi d'un erratum ou d'un corrigendum)
 - **Est complété par*** (cas d'un article suivi d'un addendum)
- **Type* F** (peut prendre la valeur **erratum** (feuille volante signalant des erreurs typographiques), **corrigendum** (article indépendant corrigeant *a posteriori* un article déjà paru), **communications** (article court ou note : pas de lien vers le plein-texte, fichier image multipage associé en un bloc), **addendum** (article en suivant immédiatement un autre, parfois inséré à la suite, souvent signé d'un auteur différent) ou **spécial** pour les cas particuliers) ; dans le cas d'un erratum ou d'un corrigendum, une relation sera insérée :
 - **Corrige**
dans le cas d'un addendum :
 - **Complète**
- **Note F** permet de signaler les anomalies ou les particularités concernant l'article, comme l'existence d'une planche hors-texte, une relation donnera l'identifiant physique et logique vers cet élément
- **Correspondance physique(cphys)** la liste des fichiers images monopages dans le système de fichiers constituant le présent article.

❖ Spécial R

- **Type (page de titre, sommaire, index, page de publicité, page blanche, ours, planche, hors-texte, spécial)**
- **Page initiale**, le contenu sera un nombre en chiffres arabes (attributs **F** : **romain, doublon, supplément, spécial**) une planche ou un hors-texte inséré entre deux pages numérotées à la suite, sera identifiée par le numéro de la page précédente, avec attribut **supplément**, une feuille volante sera déclarée en page 0, le numéro d'ordre sur cette page permettra de les distinguer
- **Page finale**, le contenu sera un nombre en chiffres arabes (attributs **F** : **romain, doublon, supplément, spécial**)
- **Numéro d'ordre sur la première page** (en général 0)
- **Note F** permet de signaler les anomalies ou les particularités, comme l'existence d'une planche hors-texte, une relation donnera l'identifiant physique et logique vers cet élément
- **Correspondance physique** (une relation vers la liste des identifiants physiques (ou nom des fichiers images monopages dans le système de fichiers) constituant le présent élément)

Les fichiers plein-texte produits pour les articles uniquement seront codés en XML selon une DTD très simple dotée d'une balise **page** et comportant l'identifiant NUMDAM. Un format plus riche (indiquant les coordonnées de chaque mot reconnu sur l'image) serait apprécié.

Ces fichiers sont codés en ISO-Latin-1. Les caractères n'appartenant pas à ce codage seront saisis à l'aide de références de caractères (p. ex. : ģ). Si les symboles mathématiques peuvent être reconnus, ils seront inclus selon leur codage Unicode, ou le nom des entités MathML associées.

B. DTD décrivant un volume physique

```
<?xml version="1.0" encoding="ISO-8859-1"?>

<!--
#
# DTD name : volphys.dtd
# Date: 18/09/2002
# Version : 1.2
# Description : Cette DTD decrit la structure des Volumes des journaux NUMDAM-->

<!ELEMENT volphys (notice, article+, special*)>
<!ENTITY % outfields "(nom, prenom, initiale?,mel?)">
<!ENTITY % pageatts 'systnum (arabe |romain) "arabe"
    pagination (normal | doublon | volant) "normal"
    typepag (normal | supplement | special | vie | preliminaire) "normal" '>
<!-- notice -->
<!ELEMENT notice (idvol, revue, serie*, tome, fascicule?, resp*, editeur, pages,
numerisation, note*)>
<!ELEMENT idvol (#PCDATA)>
<!--idvol est un identifieur qui identifie le volume avec l'identifiant du journal,
normalement il est contruit a partir des données ci dessous (acronumdam, tome etc..)-->
>
<!ELEMENT revue (issn, acronumdam)>
<!ELEMENT issn (#PCDATA)>
<!ELEMENT acronumdam (#PCDATA)>
<!--acronumdam est l'acronyme du journal-->
<!ELEMENT serie (#PCDATA)>
<!ELEMENT tome (numero, annee)>
<!ELEMENT numero (#PCDATA)>
<!--numero est un entier de n à n-->
<!ELEMENT annee (#PCDATA)>
<!ELEMENT fascicule (#PCDATA)>
<!ELEMENT resp (#PCDATA)>
<!--resp est le responsable de l'édition du volume ex actes d'un congrès)-->
<!ELEMENT editeur (#PCDATA)>
<!-- editeur est le détenteur du copyright-->
<!ELEMENT pages (#PCDATA)>
<!--pages est un entier nombre total de pages physiques-->
<!ELEMENT numerisation (idphys, datescan, infos)>
<!--numerisation ainsi que ses sous elements sont remplis par l'opérateur-->
<!ELEMENT idphys (#PCDATA)>
<!--correspondance avec le répertoire correspondant dans le système de fichiers
physique-->
<!ELEMENT datescan (#PCDATA)>
<!-- format date aaaa-mm-jj-->
<!ELEMENT infos (#PCDATA)>
<!--info concernant les caractéristiques du volume ou la numérisation, résolution, niveau
de gris, couleur, rempli de manière automatique à partir des infos fournies par le
scanneur-->
<!ELEMENT note (#PCDATA)>
<!--permet de signaler les anomalies ou particularités concernant le volume-->
<!-- /notice -->
<!-- special -->
<!ELEMENT special (pagedeb, pagefin, langue*, ordre?, note?, cphys)>
```

```

<!--special (un ou plusieurs éléments) ainsi que ses sous elements sont remplis par
l'opérateur, nous attendons surtout les éléments de type photo, planche, carte et
horstexte, nous avons laissé les autres types au cas où ils serviraient un jour-->
<!ATTLIST special
    type (pagetitre | sommaire | index | pagepub | pageblanche | ours | horstexte |
special | planche | carte | photo) #REQUIRED
>
<!ELEMENT pagedeb (#PCDATA)>
<!ATTLIST pagedeb %pageatts;
>
<!-- pagedeb et pagefin risquent souvent d'avoir la même valeur, et l'attribut pagination
risque souvent d'avoir la valeur volant-->
<!ELEMENT pagefin (#PCDATA)>
<!ATTLIST pagefin %pageatts;
>
<!ELEMENT langue (#PCDATA)>
<!ELEMENT ordre (#PCDATA)>
<!--Ordre dans la page, en général vaut 0-->
<!ELEMENT cphys (#PCDATA)>
<!--correspondance physique relation vers la liste des identifiants physiques ou noms des
fichiers monopages dans le système de fichiers constituant le présent élément)-->
<!-- /special -->
<!-- article -->
<!ELEMENT article (idart, ordreart?, pagedeb, pagefin, ordre, auteur+, contributeur*,
traducteur*, titre+, relations*, langue+, note?, biblio?, cphys?)>
<!--les éléments biblio et cphys sont remplis par l'opérateur, les autres sont déjà
générés par CMD-->
<!ATTLIST article
    type (normal | preface | postface | erratum | corrigendum | addendum |pv)
"normal"
>
<!ELEMENT idart (#PCDATA)>
<!--idart correspond a idvol plus numero d'article-->
<!ELEMENT ordreart (#PCDATA)>
<!--ordre de l'article dans la revue - pagination spéciale pour le JDEP dont chaque article
est paginé de 1 à n-->
<!ELEMENT auteur %autfields; >
<!ELEMENT nom (#PCDATA)>
<!ELEMENT prenom (#PCDATA)>
<!ELEMENT initiale (#PCDATA)>
<!ELEMENT mel (#PCDATA)>
<!ELEMENT contributeur %autfields; >
<!ATTLIST contributeur
    role (traducteur | redacteur ) #IMPLIED
>
<!ELEMENT traducteur %autfields; >
<!ELEMENT titre (#PCDATA)>
<!--xml:lang est un attribut prédéfini codes en 2 cars, orig est le titre dans la langue
originel-->
<!ATTLIST titre
    xml:lang CDATA #IMPLIED
    orig CDATA #IMPLIED
>
<!ELEMENT relations (#PCDATA)> <!-- relation vers l'idnumdam de l'article-->
<!-- obligatoire si l'article a un type erratum corrigendum ou addendum, l'attribut doit
être mis en conséquence?-->
<!ATTLIST relations

```

```
type (corrige | estcorrige | complete | estcomplete) #REQUIRED
>
<!ELEMENT biblio (bibitem+)><!-- tout ce qui suit est rempli par l'opérateur-->
<!ELEMENT bibitem ( #PCDATA | bauteur | btitre | bannee | bpagedeb)*>
<!--ELEMENT bauteur %autfields;-->
<!ELEMENT bauteur (#PCDATA)>
<!ELEMENT btitre (#PCDATA)>
<!ELEMENT bannee (#PCDATA)>
<!ELEMENT bpagedeb (#PCDATA)>
<!-- /article -->
```

Annexe 4 : traitement des bibliographies

Un premier niveau de balisage est obligatoire pour les bibliographies, à savoir :

❖ Article R

➤ Biblio

- **Bibitem R** (contenant chaque référence bibliographique contenue dans l'article en question), à l'intérieur de laquelle on balisera l'information suivante :
 - **auteur R**
 - **titre**
 - **année de publication**
 - **première page F**

DTD correspondante :

```
<!ELEMENT biblio (bibitem+)>
<!ELEMENT bibitem ( #PCDATA | bauteur | btitre | bannee | bpagedeb)*>
<!--ELEMENT bauteur %autfields;-->
<!ELEMENT bauteur (#PCDATA)>
<!ELEMENT btitre (#PCDATA)>
<!ELEMENT bannee (#PCDATA)>
<!ELEMENT bpagedeb (#PCDATA)>
```

Un deuxième niveau, optionnel, sera proposé par le soumissionnaire :

❖ Article R

➤ Biblio

- **Bibitem R** (contenant chaque référence bibliographique contenue dans l'article en question), à l'intérieur de laquelle on balisera l'information suivante :
 - **Auteur R**
 - ◆ **prénom**
 - ◆ **nom**
 - **titre**
 - **sous-titre F**
 - **nom de la revue F**
 - **série F**
 - **tome F**
 - **fascicule F**
 - **numéro d'édition F**
 - **éditeur F**
 - **lieu d'édition F**
 - **année de publication**
 - **page initiale**
 - **page finale F**
 - **collection**
 - **URI F**

Exemple de bibliographie :

BIBLIOGRAPHIE

- [1] N. BOURBAKI, *Algèbre*, chap. II (Algèbre Linéaire), 3^e édition, Hermann, Paris, 1962.
- [2] N. BOURBAKI, *Algèbre*, Chap. III (Algèbre Multilinéaire), 2^e édition, Hermann, Paris, 1958.
- [3] N. BOURBAKI, *Algèbre Commutative*, chap. I et II, Hermann, Paris, 1961.
- [4] N. BOURBAKI, *Algèbre Commutative*, chap. VI, à paraître.
- [5] P. CARTIER, Questions de rationalité des diviseurs en Géométrie Algébrique, Appendice, *Bull. Soc. Math. France* 86 (1958), 177-251.
- [6] C. CHEVALLEY, *Théorie des Groupes de Lie*, tome II (Groupes Algébriques), Hermann, Paris, 1951.
- [7] C. CHEVALLEY, The construction and study of certain important algebras, *The Mathematical Society of Japan*, Tokyo, 1955.
- [8] C. CHEVALLEY, Fundamental concepts of algebra, *Academic Press Inc.*, New York, 1956.
- [9] A. DOLD, Les foncteurs dérivés d'un foncteur non-additif, *Séminaire Bourbaki*, 1958-1959, Exposé 170.
- [10] A. DOLD, Homology of symmetric products and other functors of complexes, *Ann. of Math.* 68 (1958), 54-80.
- [11] A. DOLD and D. PUPPE, Non-additive functors, their derived functors and the suspension homomorphism, *Proc. Nat. Ac. Sc. USA* 44 (1958), 1065-1068.
- [12] A. DOLD und D. PUPPE, Homologie nicht-additiver Funktoren, Anwendungen, *Annales Inst. Fourier*, 11 (1961), 201-312.
- [13] G. GLAESER, Théorie intrinsèque des polynômes et dualité, *Bull. Sc. Math.*, 85 (1961), 17-28.
- [14] A. GROTHENDIECK et J. DIEUDONNÉ, *Éléments de Géométrie Algébrique I*, *Publications Math. de l'IHES* n. 4, Paris, 1960.

Extraits :

[1] N. BOURBAKI, *Algèbre*, chap. II (Algèbre Linéaire), 3^e édition, Hermann, Paris, 1962.

[14] A. GROTHENDIECK et J. DIEUDONNÉ, *Éléments de Géométrie Algébrique I*, *Publications Math. de l'IHES* n. 4, Paris, 1960.

Balitage XML correspondant :

1. Balitage obligatoire

```
<biblio>
<bibitem>[1] <bauteur>N. BOURBAKI</bauteur>, <btitre>Algèbre</btitre>, chap. II
(Algèbre Linéaire), 3e édition, Hermann, Paris, <bannee>1962</bannee>.</bibitem>
<bibitem> .... </bibitem>
<bibitem> .... </bibitem>
<bibitem>[14] <bauteur>A. GROTHENDIECK</bauteur> et <bauteur>J.
DIEUDONNÉ</bauteur>,
<btitre>Éléments de Géométrie I</btitre>, Publications Math. de l'IHES n. 4, Paris,
<bannee>1960</bannee>.</bibitem>
<bibitem> .... </bibitem>
<bibitem> .... </bibitem>
</biblio>
```

2. Balisage optionnel

```
<biblio>
<bibitem>[1] <bauteur><bprenom>N.</bprenom>
<bnom>BOURBAKI</bnom></bauteur>, <btitre>Algèbre</btitre>, <bsoustitre>chap.
II (Algèbre Linéaire)</bsoustitre>, <bedition>3e édition</bedition>,
<bediteur>Hermann</bediteur>, <blieued>Paris</bleiued>,
<bannee>1962</bannee>.</bibitem>
<bibitem> .... </bibitem>
<bibitem> .... </bibitem>
<bibitem>[14] <bauteur><bprenom>A.</bprenom>
<bnom>GROTHENDIECK</bnom></bauteur> et <bauteur><bprenom>J.</bprenom>
<bnom>DIEUDONNÉ</bnom></bauteur>, <btitre>Éléments de Géométrie I</btitre>,
<brevue>Publications Math. de l'IHES</brevue> n. <bvol>I</bvol>,
<blieued>Paris</bleiued>, <bannee>1960</bannee>.</bibitem>
<bibitem> .... </bibitem>
<bibitem> .... </bibitem>
</biblio>
```

Annexe 5 : plein-texte d'un article

DTD structurant le plein-texte d'un article

```
<?xml version="1.0" encoding="ISO-8859-1"?>
< !—
#
# DTD name : article.dtd
# Description : cette DTD décrit la structure des articles XML de NUMDAM

<!-- article -->
<!ELEMENT article (body)>
<!ATTLIST article
    idart CDATA #REQUIRED>

<!ELEMENT body (page+)>

<!ELEMENT page (#PCDATA)>
<!ATTLIST page
    numero CDATA #REQUIRED>

<!-- /article -->
```

Exemple :

```
<article idart="AIF_1986__36_2_1_0">
<body>
<page numero="1"> Ann. Inst. Fourier, Grenoble
36, 2 (1986), 1-25.
AUTOMATES FINIS ET ENSEMBLES NORMAUX
par Christian MAUDUIT
1. Introduction.
Rappelons que l'on appelle ensemble normal associé à une suite  $u = (u_n)$ 
l'ensemble des nombres réels  $R$  tels que la suite  $(M_n)_{n \in \mathbb{N}}$  soit équirépartie
modulo  $u_n$ .
Le but de cette étude est de déterminer une condition nécessaire et suffisante
pour que l'ensemble normal associé à une suite d'entiers  $U = (u_n)$ 
' reconnaissable par un automate fini soit exactement  $R \setminus \mathbb{Q}$ 
(la définition précise de la reconnaissabilité sera rappelée ultérieurement).
D'après le théorème de Weyl (cf [15] ou [20]) le problème revient à étudier le
comportement de la somme  $e(U_n)$  et de déterminer à  $n \in \mathbb{N}$  1 quelles
conditions on a, pour tout  $\alpha$ , irrationnel  $\lim_{n \rightarrow \infty} e(U_n \cdot \alpha) = 0$  (en posant  $e(x) = e^{2\pi i x}$ ).
Résultats antérieurs. — En 1980, Kamae, dans un article non publié (cf [7]),
avait donné une ébauche de démonstration dans le cas particulier où  $u_n = O(n)$ 
(par exemple pour la suite de Morse (cf exemple 2)).
</page><page numero="2">...</page>
</body>
</article>
```

Annexe 6 : cadre de réponse

Les opérateurs devront fournir les documents suivants :

Présentation de la société :

- Organigramme de la société
- Chiffre d'affaire 1999, 2000 et 2001
 - Global
 - Pour l'activité de numérisation
- Implantations géographiques

Effectif en personnel :

- Effectif global
- Part des effectifs permanents spécialisés dans l'activité de conversion de données
 - Pour le traitement des données
 - Pour le personnel d'encadrement

Curriculum vitæ de l'équipe proposée

- CV du chef de projet responsable de l'opération
- CV des membres de l'équipe

Localisation géographique des différents traitements