



**PRÉFET
DE LA RÉGION
OCCITANIE**

*Liberté
Égalité
Fraternité*

**Direction régionale de l'environnement,
de l'aménagement et du logement**

CAHIER

des

CLAUSES

TECHNIQUES

PARTICULIERES

relatif à l'appel d'offre pour

**la sauvegarde des données anciennes pluviométriques
et hydrométriques de la DREAL Occitanie**

Table des matières

1 – Présentation générale du marché.....	3
2 – Contexte et enjeux.....	4
4 - Description détaillée des prestations de l'étape d'initialisation.....	7
4.1 - Procédure d'étalonnage et de test initial pour le scannage.....	7
4.2 - Procédure d'étalonnage et de test initial pour la numérisation.....	8
5 - Description détaillée des prestations de l'étape de sauvegarde.....	8
5.1 – Phase 1 : scannage.....	9
5.2 - Phase 2 : numérisation.....	11
6 – Réunions et livrables.....	14
6.1 – Réunions.....	14
6.2 – Livrables.....	14
Annexe 1 : acronymes et définitions.....	16
Annexe 2: structure de fichiers d'entrée et de sortie, nomenclature.....	17
1 – Fichier d'inventaire priorisé.....	17
2 – Fichier de métadonnées.....	17
3 – Nomenclature.....	17
Annexe 3: exemples de documents à scanner.....	19
1 – Pluviogramme.....	19
2 – Liminigramme.....	20
3 – Tableau.....	22
Annexe 4: Fichier de rendu pour la numérisation des tableaux.....	23

1 – Présentation générale du marché

L'objet du marché est de sauvegarder (c'est-à-dire de scanner, de numériser, et de documenter) plusieurs ensembles de données pluviométriques et hydrométriques actuellement conservées sous forme papier dans les services de la DREAL Occitanie, et qu'on appellera « gisements de données ». La nomenclature des divers termes utilisés figure en **annexe 1**.

La DREAL Occitanie est désignée Maître d'Ouvrage (MO). Elle organise la procédure de passation, conclut le marché, dans le cadre duquel elle pourra émettre des bons de commande tels que définis à l'article 1.4 du CCAP, et suivre l'exécution du marché.

Le marché est conclu pour une durée de 12 mois à compter de la date de notification de l'ordre de service déclenchant le commencement d'exécution, reconductible 3 fois pour une durée de 12 mois.

Il comprend :

- une étape d'initialisation consacrée à la présentation du marché, la formation du titulaire au logiciel de numérisation NUNIEAU et le test sur un échantillon réduit de données pour le scannage et la numérisation. Ses caractéristiques et durées sont indiquées au paragraphe 4.
- une étape de sauvegarde qui consiste en un traitement (scannage et/ou numérisation) des gisements de données au fur et à mesure de la survenance des besoins.

Pour chaque gisement de données, les principales phases de la prestation sont représentées sur la **figure 1**.

La 1^{ère} phase consiste en un **scannage** des divers documents sous forme papier. Le titulaire transformera les données archivées en images scannées (au format .pdf et .tiff), en s'aidant d'une part des métadonnées associées et d'autre part de l'inventaire priorisé établi par la MO. Il établira en outre un rapport d'exécution du scannage (art. 5.1.8).

Lors de la 2^{ème} phase, dite de **numérisation**, le titulaire repartira des images scannées pour les transformer en fichiers numériques. Il établira en outre un rapport d'exécution de la numérisation (art. 5.2.7).

Il est à noter que certains documents, déjà scannés, seront fournis uniquement pour numérisation.

Par ailleurs, certains documents ne pourront être numérisés qu'à un deuxième tour ou pas du tout.

La prestation conduira à utiliser un très grand nombre de données et à produire un nombre comparable de fichiers informatiques. De ce fait, les documents associés au traitement des données (inventaire priorisé, rapports d'exécution) revêtent une importance particulière. En particulier, ce sont eux qui permettront :

- de définir le déroulement des traitements
- de préciser la nomenclature des fichiers
- de pouvoir documenter les différentes phases de la prestation.

La structure de ces fichiers est précisée à l'annexe 2.

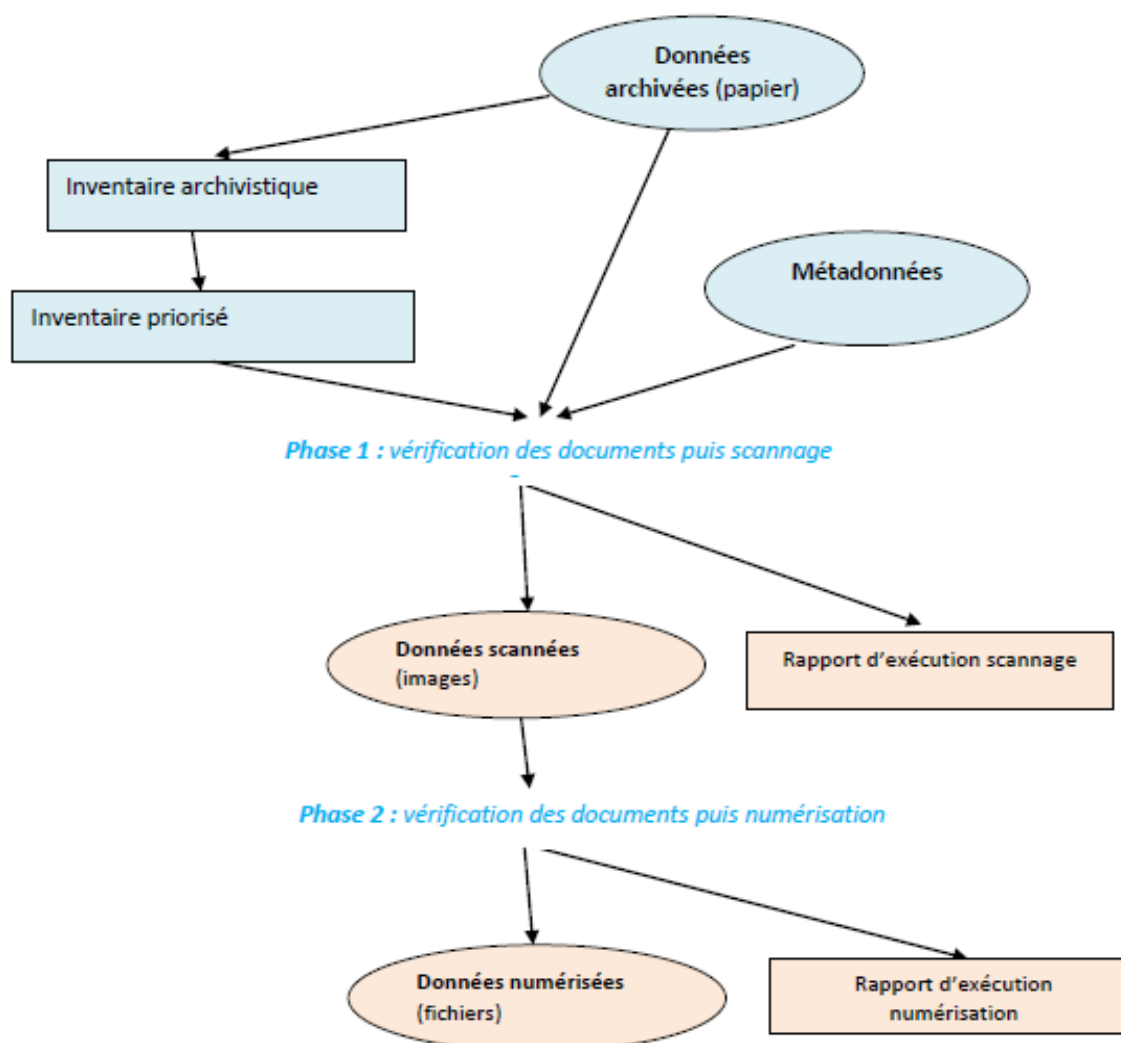


Figure 1: Principales phases de l'étape de sauvegarde (les documents produits par la MO sont colorés en bleu, les documents à produire par le titulaire sont colorés en beige)

2 – Contexte et enjeux

La Direction régionale de l'environnement, de l'aménagement et du logement (DREAL) Occitanie est un service déconcentré du ministère de la Transition écologique et de la Cohésion des territoires. Elle est placée sous l'autorité du préfet de région pour une grande partie de ses missions et sous l'autorité des préfètes et préfets de département pour certaines missions telles que la prévention des risques, les sites inscrits et classés.

La DREAL Occitanie héberge deux services de prévision des crues et d'hydrométrie, les SPC Méditerranée Ouest et SPC Garonne Tarn Lot, l'un basé à Carcassonne, l'autre à Toulouse. Les

Unité d'Hydrométrie gèrent l'installation et le fonctionnement des points de mesure pluviométriques et hydrométriques de l'Etat.

Les observations pluviométriques et hydrométriques méritent une attention particulière car elles constituent la principale donnée d'entrée pour les outils de la surveillance et la prévision des crues. La valorisation, de toutes les données disponibles, en temps réel ou en temps différé, constitue un point-clé dans ce domaine.

Un effort important a été mené depuis les années 1990 pour mettre en place des bases de données hydrométéorologiques. Elles permettent d'accéder facilement aux données contemporaines, ce qui satisfait déjà un grand nombre d'utilisateurs potentiels. Néanmoins, la profondeur d'archive reste actuellement limitée, et des données anciennes, disponibles sous diverses formes (graphiques ou tableau papier, format informatique spécifique), ne sont pas encore incluses dans ces bases de données.

Ces données anciennes sont particulièrement précieuses car :

- elles permettent d'améliorer l'estimation de la fréquence des phénomènes extrêmes sur un lieu donné, par l'allongement des séries temporelles associées aux postes de mesure ;
- elles permettent de recenser l'existence d'événements extrêmes, mais anciens, qui se sont déroulés dans le passé et donc susceptibles de se reproduire (exemple des crues observées sur la Garonne marmandaise en février 2021 et 2026) ;
- de façon plus globale, elles constituent un patrimoine unique dont l'utilité s'étend à d'autres domaines, en particulier les études climatiques.

Leur devenir reste précaire. Les risques de détérioration, d'effacement de l'encre, d'oubli et de perte ne sont pas négligeables, notamment à l'occasion du transfert des organismes qui les hébergent. La DREAL Occitanie souhaite contribuer à leur sauvegarde.

La DREAL Occitanie dispose d'un gisement d'archives hydrométriques et pluviométriques conséquent, qui représente plus de 100 ml. Ces archives ont fait l'objet d'un recollement en 2022 et 2023. Une cinquantaine de stations, dont un grand nombre sont encore actives aujourd'hui, présentent un intérêt, quant à la numérisation de leurs données et au reversement de leur chroniques et métadonnées dans la base nationale de données hydrométriques (www.hydro.eaufrance.fr).

3 – Description détaillée des documents

Les documents à traiter sont de différents types :

- documents sous forme de graphiques : pluviogrammes ou limnigrammes ;
- documents sous forme de registres : tableaux de relevés de pluie ou de hauteur d'eau, reliés ou pas ;
- d'autres documents, permettant une description et une exploitation des graphiques (appelées « métadonnées d'exploitation »), seront également à traiter ponctuellement.
- registres déjà scannés : il s'agit d'une minorité de documents.

Il s'agit de documents datant du 20^{ème} siècle, de la 2^{ème} moitié du 19^{ème} siècle voire de la 1^{ère} moitié.

Une description plus précise de ces données est fournie à **l'annexe 3**.

A ce stade, l'ensemble des données à traiter n'est pas connu avec précision. Elles le seront au moment de l'émission de chacun des bons de commande.

Toutefois, un premier recensement précis a été effectué pour certaines stations, dont le scan et la numérisation des données a été jugée prioritaire. Le tableau suivant présente à titre indicatif la catégorisation de quelques documents à traiter, en distinguant les documents reliés des documents non reliés, et les formats.

Station	Période	nb tableaux	nb graphiques
Saint-Béat	1857-1989	1459	0
Chaum	1857-2000	1504	712
Toulouse Embouchure	1857-1941	985	0
Valentine	1857-1990	1007	2581
Cazères	1956-1983	67	744

Station	nb feuilles reliées	nb feuilles		
		non reliées	nb feuilles <A3	nb feuilles >A3
Saint-Béat	161	1298	1332	132
Chaum	299	1920	1304	921
Toulouse Embouchure	277	712	874	116
Valentine	391	3235	2855	771
Cazères	50	765	124	692

De façon plus générale, les tableaux qui suivent représentent un aperçu général de leurs caractéristiques, fourni de façon à permettre aux candidats de cadrer leur proposition.

Au moins 3 gisements de données seront à traiter au cours du marché.

Tableau 1 : Données scientifiques : pluviogrammes	
Nombre approximatif	De 100 à 3200 documents par gisement
Stockage	Variable : parfois à plat, souvent en rouleaux attachés par des élastiques, autres modes possibles.
Etat	Bon état en général. Néanmoins, des documents peuvent présenter des traces d'usure, de déchirement au bord des feuilles.
Lisibilité	Généralement bonne, mais certaines courbes peuvent être peu visibles (l'encre s'effaçant au fil du temps, ou des tâches apparaissant)
Support papier	Recto, monochrome ou couleur, sur feuilles lignées. Des informations complémentaires sont souvent inscrites en superposition.
Format	Très divers. Une partie inférieure à A3, une partie sous forme de rouleaux (hauteur inférieure à A4) longueur jusqu'à plusieurs mètres
Classement	Un seul lieu de stockage pour un gisement de données.

Tableau 2 : Données scientifiques : limnigrammes	
Nombre approximatif	De 100 à 3200 documents par gisement
Stockage, conditionnement	Variable : parfois à plat, souvent en rouleaux attachés par des élastiques, autres modes possibles.
Etat de conservation	Généralement plutôt bon, mais certaines courbes peuvent être peu visibles (l'encre s'effaçant au fil du temps, ou des tâches apparaissant)
Lisibilité	Assez bonne à médiocre
Support papier	Recto, monochrome ou couleur, sur feuilles lignées. Des informations complémentaires sont parfois inscrites en superposition.
Format	Une partie inférieure à A3, une partie sous forme de rouleaux (hauteur inférieure à A4) longueur jusqu'à plusieurs mètres
Classement	Un seul lieu de stockage pour un gisement de données.
Tableaux 3 : Données scientifiques : tableaux de données	

Nombre approximatif	De 501 à 1800 documents par gisement
Stockage, conditionnement	Feuilles volantes ou reliées (par 12 généralement) insérées dans des chemises, registres
Etat de conservation	Très variable (nombreux documents du 19^{ème} siècle)
Lisibilité	Souvent bonne
Support papier	variable
Format	Majoritairement inférieur à A3
Classement	Un seul lieu de stockage pour un gisement de données.

La DREAL Occitanie souhaite numériser les données de 10 à 20 stations par an.

4 - Description détaillée des prestations de l'étape d'initialisation

Cette étape, pilotée par la DREAL Occitanie consiste en un test préalable et un étalonnage des prestations de scannage et de numérisation en amont de l'étape de sauvegarde.

4.1 - Procédure d'étalonnage et de test initial pour le scannage

Le titulaire procédera au scannage d'un jeu d'essai défini par la MO et constitué de et 20 limnigrammes ainsi que de 60 tableaux, accompagnés de leurs éventuels documents annexes. Cette procédure, destinée à garantir la qualité du scannage, fait partie intégrante de la prestation. Le titulaire procédera à la vérification de la conformité des documents scannés avec les exigences techniques (5.1.7) du CCTP et rédigera un procès-verbal rendant compte du résultat de ce contrôle. Le délai pour ce test sera de deux semaines semaine.

A l'issue de la phase d'étalonnage, les fichiers et les documents du jeu d'essai seront fournis à la MO pour comparaison avec les documents originaux, qui disposera d'un mois pour effectuer la vérification. Le procès-verbal du contrôle effectué par le titulaire sera joint.

Si le résultat des essais est satisfaisant pour les deux parties, les paramètres des réglages effectués pour les réaliser seront considérés comme valeurs étalons pour la suite de la prestation.

Si le résultat des tests n'est pas satisfaisant pour la MO, un deuxième jeu d'essai sera préparé par la MO, et réalisé dans les mêmes conditions.

Si le résultat du 2ème test n'est pas satisfaisant, la MO se réserve le droit de résilier le marché. Le titulaire devra alors restituer à la MO l'intégralité des documents papier et scannés dans un délai de 10 jours ouvrés.

A l'issue de la phase d'étalonnage et de test initial pour le scannage, si celle-ci est positive, une réunion technique physique ou téléphonique sera organisée pour lancer la procédure d'étalonnage et de test initial pour la numérisation.

4.2 - Procédure d'étalonnage et de test initial pour la numérisation

Logiciel de numérisation

Pour les données de graphique (pluviogrammes, hydrogrammes), la DREAL Occitanie souhaite numériser les données à partir des images obtenues dans la phase précédente (scans). **Le titulaire pourra avoir recours au logiciel gratuit et libre d'utilisation NUNIEAU**, développé par le Cerema (Direction Territoriale Méditerranée) pour la numérisation de limnigrammes et de pluviogrammes papiers. On trouvera le détail de la méthode sur le lien : http://www.cete-mediterranee.fr/fr/rubrique.php3?id_rubrique=113.

Afin de familiariser le titulaire à l'utilisation du logiciel, une formation de 3 jours sera dispensée par le Cerema dans les locaux de la DREAL Occitanie à Toulouse.

Par la suite, un dialogue régulier sera indispensable entre le titulaire et le MO. Le titulaire pourra poser toutes les questions nécessaires au bon déroulement des étapes de « calage » puis « numérisation », à l'occasion de la phase de test, et plus tard en cas de besoin.

Variante :

Toutefois, si le titulaire dispose déjà d'un autre logiciel de numérisation auquel il serait déjà formé et permettant d'arriver au même résultat, il pourra le proposer dans son offre. La validation de son utilisation se fera à l'issue de la phase de test. Le MO se réservera alors le droit d'imposer le logiciel NUNIEAU s'il estime que les performances du logiciel proposé ne sont pas suffisantes.

Dans cette phase initiale de test, on procédera à un test sur la numérisation en continuité du test déjà mis en œuvre pour le scannage, à partir du même jeu d'essai. Le test interviendra donc après la formation du titulaire au logiciel NUNIEAU, si c'est le logiciel retenu.

Le titulaire procédera à la rédaction des rapports d'exécution associées. Le test comprend également la mise à jour du fichier d'inventaire. Ce jeu d'essai, destiné à garantir la qualité de la numérisation, fait partie intégrante de la prestation. Le titulaire disposera de deux semaines pour le mettre en œuvre.

A l'issue de la phase test, les fichiers image et les résultats de l'essai seront renvoyés à la MO pour comparaison. Le procès-verbal du contrôle effectué par le titulaire sera joint. La MO disposera d'un mois pour effectuer la vérification.

Si le résultat des essais est satisfaisant pour les deux parties, le protocole de numérisation mis en place sera considéré comme opérationnel pour la suite de la prestation. Si le résultat des tests n'est pas satisfaisant, un deuxième jeu d'essai sera préparé par la MO.

Si le résultat du deuxième test n'est pas satisfaisant, la MO se réserve le droit de résilier le marché.

A l'issue de la phase de test, si celle-ci est positive, une réunion de bilan physique ou visio-téléphonique sera organisée.

5 - Description détaillée des prestations de l'étape de sauvegarde

Une réunion de lancement de l'étape de sauvegarde sera systématiquement programmée. Elle permettra au MO de présenter l'inventaire priorisé (quantité, type de documents, métadonnées associées) et les conditions de réalisation de la prestation. Le devis relatif au bon de commande ne sera établi que postérieurement à cette réunion de lancement. Cette réunion de lancement pourra être réalisée en visio-conférence.

5.1 – Phase 1 : scannage

5.1.1 - Liste des documents à scanner

Le fichier d'inventaire priorisé et le fichier de métadonnées seront fournis par la MO au titulaire au début de chaque prestation. On en trouvera une description à l'annexe 2.

Si des documents techniques ou annexes sont fournis en plus ou en moins, l'inventaire devra être amendé en ce sens dans le cadre de la vérification préalable à tout scannage (cf. 5.1.5) par le titulaire. Cet inventaire numérique servira pour le suivi et la traçabilité de l'action de scannage et les amendements éventuels seront complétés au fur et à mesure de l'avancement de la tâche.

5.1.2 - La poussière

Des opérations ponctuelles de dépoussiérage des documents pourront être nécessaires avant scannage. En effet, les documents disposés au dessus de piles et les extrémités des rouleaux conservés aux archives présentent des traces de poussière non compatibles avec un scanner à défilement. Le cas échéant, un dépoussiérage préalable devra être effectué par le titulaire.

Toutefois, il est précisé que l'ensemble des documents à scanner qui seront fournis par le prestataire a fait l'objet d'une prestation préalable de contrôle sanitaire de support et, le cas échéant, de dépoussiérage et de décontamination.

5.1.3 - Documents abîmés

Des opérations de préparation des documents de type pluviogrammes ou limnigrammes pourront également être nécessaires. Pour certains, les tracés couvrent la feuille jusqu'aux extrémités qui sont parfois altérées. Un ruban adhésif (type repositionnable et si possible neutre (sans acide) devra être collé au niveau de l'extrémité afin de pouvoir bien scanner l'intégralité du document. Dans le cas de documents très abîmés ou très sales, les documents (en intégralité ou leurs extrémités) pourront être positionnés entre deux films plastiques. La réunion de lancement permettra de déterminer l'étendue et les conditions de réalisation de ces travaux préparatoires.

5.1.4 - Lieu et conditions d'exécution de la prestation

Dans certains cas la sortie des documents peut ne pas être autorisée par le service qui en assure la conservation. Le titulaire devra alors mettre en œuvre les moyens nécessaires à la réalisation de la prestation de scannage dans les locaux du MO, qui de son côté assurera les meilleures conditions pour la réalisation de cette prestation. La personne publique indiquera le lieu précis de la prestation, les horaires d'accès aux locaux, les modalités d'accès, la manière dont les documents sont acheminés depuis les magasins de stockage vers le lieu de la prestation, les équipements pouvant être fournis au titulaire.

Dans les autres cas, le titulaire exécutera la mission de scannage dans sa propre structure. Pour cela, un prêt de documents sera mis en place et le titulaire viendra récupérer les documents au fur et à mesure des commandes. Les méthodes utilisées seront laissées à l'appréciation du titulaire mais devront veiller aux prescriptions définies au 5.1.5.

Les conditions seront déterminées d'un commun accord entre le titulaire et le MO lors de la réunion de lancement de la prestation.

5.1.5 - Prise en charge des documents

Dans le cas où la prestation est réalisée dans des locaux désignés par le titulaire, des opérations de prise en charge, transport et stockage des documents seront nécessaires. Un planning de prise en charge sera établi en concertation avec le MO. Il sera validé en commun entre le titulaire et le MO.

Le titulaire assurera la manutention, le transport et le stockage des documents à scanner. Les précautions nécessaires seront prises pour protéger les documents pendant leur prise en charge et leur transport : emballage et protection contre la pluie par exemple.

Aucun document ne sera autorisé à sortir du territoire français.

La prise en charge des documents se fera dans les locaux du MO à l'adresse désignée dans les bons de commande, dans le respect du règlement intérieur et des règles de sécurité en vigueur. Le titulaire sera responsable des documents depuis leur prise en charge jusqu'à leur retour au MO. Durant le transport et le stockage des documents, la température devra être régulée à environ 20°C, en évitant tout choc thermique violent et toute exposition directe au soleil.

Pour la durée de leur utilisation par le titulaire, les documents seront stockés dans un lieu garantissant des conditions correctes de sécurité (local fermé à clé) et respectant les règles concernant la lutte contre l'incendie et les inondations. Les coordonnées de ce lieu seront précisées dans le planning de charge. La MO se réserve la possibilité de vérifier sur place les conditions de stockage des documents et de réalisation des prestations.

Avant de réaliser le scannage des documents qui lui ont été confiés, le titulaire procède à une vérification préalable quantitative et qualitative :

- Le titulaire vérifiera au moment de la prise en charge, l'adéquation entre les informations figurant sur l'inventaire priorisé et les documents qui lui sont confiés, à charge pour lui de demander tout complément d'information auprès du commanditaire pour assurer la prestation dans les meilleures conditions. La signature qu'il apposera sur le bordereau d'envoi vaudra acceptation de l'ensemble.
- Le titulaire signalera toute anomalie et tout document qu'il considérera comme impropre au scannage ou à la numérisation avec le motif précis du rejet. Il appartiendra au MO de décider si le document concerné doit être retiré du lot ou non. Les documents seront réputés en bon état, hormis ceux signalés comme tels. Ils devront être restitués dans un état identique. Le titulaire préviendra le MO de la date de restitution des documents avec un préavis minimum de 5 jours ouvrés.

5.1.6 – Nomenclature

Comme précisé dans la présentation générale du marché, le respect de la nomenclature est un élément essentiel du projet.

Les règles de nomenclature devront s'appliquer successivement :

- A l'arborescence et au nommage des répertoires.
- Aux fichiers eux-mêmes

Un exemple de nommage est proposé à l'annexe 2. Le titulaire pourra l'amender et en proposer une amélioration lors de la réunion de lancement du marché.

Les scans des documents annexes considérés comme métadonnées d'exploitation seront enregistrés dans les répertoires des documents techniques correspondants. S'il y a plus d'un scan, un dossier sera créé avec la nomenclature adéquate pour les rassembler.

5.1.7 - Exigences techniques

Les exigences sur les documents scannés sont les suivantes :

- images sous deux formats : .tif et .pdf. La compression de sauvegarde du format tif doit se faire sans perte d'information et permettre une réutilisation avec le logiciel de numérisation Nunieau (art. 4.2).
- résolution : 300dpi pour les pdf, 300 dpi de préférence pour les documents graphiques, en lien avec la taille du document. Par exemple, une feuille A3 sera en 300 dpi, une planche A0 sera plutôt en 100 ou 200 dpi.
- mode couleur (RGB) pour les documents en couleurs, niveau de gris pour le noir et blanc;
- nomenclature imposée (art. 5.1.6).

Le titulaire devra veiller à ce que le scannage n'induisse pas de distorsion dans l'image produite (en particulier pour les rouleaux de taille supérieure à A3). Les dimensions de l'image scannée devront être identiques à celles du graphique d'origine. Le carroyage de l'image scannée devra être fidèle à celui d'origine et ne pas présenter de courbure liée au scannage. Les fichiers images seront orientés comme le document original, dans le sens de lecture de la légende ou du titre.

5.1.8 – Exécution de la prestation de scannage

L'exécution sera formalisée par des «rapports d'exécution », le titulaire devra les renseigner pour assurer le suivi et la traçabilité de son travail. Ils devront en particulier renseigner :

- L'état général d'avancement des tâches sous forme de tableaux
- Pour chaque feuille traitée, inclure les commentaires nécessaires (par exemple lorsque l'image aura nécessité un post-traitement).

Les rapports d'exécution devront donc être accessibles en ligne par la MO, et remis à jour au moins au pas de temps mensuel.

A la fin de la prestation, ils devront être inclus dans le dossier de livrables.

5.1.9 - Procédure de vérification finale du scannage par le MO

Les opérations de vérification s'attacheront à comparer les documents scannés avec les originaux et à valider la conformité des livrables avec les exigences du CCTP. Le délai maximal dont dispose le MO pour procéder aux opérations de vérification sera précisé dans chaque bon de commande. La méthode utilisée par le MO est décrite par la norme NF ISO 2859-1. On procédera par double contrôle par échantillon. Le NQA (Niveau de Qualité Acceptable) est fixé à 2,5% (il pourra être modulé dans chaque bon de commande en fonction de l'état du gisement). A partir d'un nombre d'anomalies constaté supérieur au seuil déduit du volume de l'échantillon, la prestation sera rejetée. Le titulaire est tenu de refaire la prestation à ses frais dans un délai de 1 mois.

Un procès-verbal sera établi par le MO à l'issue des opérations de vérification.

5.2 - Phase 2 : numérisation

5.2.1 - Numérisation des pluviogrammes et limnigrammes

Dans ce paragraphe, les pluviogrammes et limnigrammes seront désignés comme « graphiques ». Leur spécificité de traitement est décrite en annexe 3.

Assemblage

La plupart des graphiques sont des enregistrements de quelques jours à une semaine. L'ensemble des séries d'un même graphique issues de la numérisation seront à assembler pour produire une série temporelle continue sur la période totale d'enregistrement, selon la demande spécifiée dans le fichier d'inventaire. Afin de garantir des fichiers de taille gérables par des éditeurs de texte, le MO se réserve le droit de spécifier, préalablement à l'édition de la commande, une durée d'assemblage qui est inférieure à la durée totale des relevés à numériser.

Cette étape se fait dans NUNIEAU (ou logiciel équivalent) et permet d'évaluer la continuité de la série temporelle.

Mentions manuscrites sur les limnigrammes

Concernant les graphiques de type limnigrammes, ceux-ci sont quasi-systématiquement accompagnés de mentions manuscrites, à l'initiative de l'observateur de crue qui relevait ces rouleaux.

Parmi ces mentions manuscrites figure quasi-systématiquement le relevé visuel, le jour du changement du rouleau, de la cote à l'échelle.

Le renseignement de cette information est capitale afin de pouvoir porter un regard critique sur le décalage éventuel de la mesure enregistrée sur le graphique.

La numérisation de ces valeurs sera réalisée dans un fichier .csv (date/heure et valeurs) qui sera livré en plus de celui contenant la chronique numérisée (ou dans un onglet spécifique de ce fichier).

D'autres mentions peuvent, de façon exceptionnelles, être reportées, comme des commentaires sur d'éventuels problèmes de mesure, la valeur d'un débit jaugé, la valeur d'un zéro d'échelle, des références liées au poste de mesure.

Ces informations seront également reportées dans un fichier .csv qui sera lié au fichier de la chronique numérisée. (la date de ce commentaire sera reportée dans une colonne dédiée)

Ces métadonnées constituent les métadonnées d'exploitation.

Afin de garantir un rendu qui représente également les micro-variations, **seront numérisés environ 12 points par heure.**

Les problèmes constituant une gêne pour la numérisation sont voisins de ceux relatifs aux pluviogrammes, à savoir le retournement du tracé lorsque le stylet atteignait les bords inférieurs ou supérieurs du papier. Le prestataire sera formé à la gestion de ces retournements lors de la formation au logiciel NUNIEAU.

Fichiers et formats attendus

Le titulaire devra fournir les résultats intermédiaires et définitifs de la numérisation des graphiques.

Les résultats intermédiaires de la numérisation sont les fichiers de calage et les fichiers bruts en sortie de logiciel propres à la numérisation de chaque courbe. Ces fichiers intermédiaires permettent la traçabilité du travail effectué.

Les résultats définitifs sont les fichiers produits sur un ensemble de graphiques comprenant la période définie. La MO demande que les fichiers numériques définitifs constitués aient le format de sortie du logiciel NUNIEAU.

5.2.2- Numérisation des tableaux

Pour les tableaux, il s'agit de reporter les valeurs notées sur les relevés papier dans un fichier texte. Le MO n'impose pas de méthode particulière pour mener à bien cette prestation.

Toutefois, toutes les valeurs présentes dans le tableau devront avoir été reportées dans le fichier texte.

Pour cela, il sera demandé de respecter un formalisme (en tête, champs remplis).

Notamment, afin de faciliter le travail de contrôle (à la fois par le prestataire mais aussi par le MO) , il sera fourni au prestataire un modèle de fichier de remplissage (fichier .ods) qui permet de réaliser un contrôle visuel des données saisies.

Chaque fichier constitué sera limité à une décennie de données, cela afin de faciliter le contrôle par le MO.

Un exemple de ce fichier est présent en annexe 4.

Ce formalisme de rendu à respecter pourra être adapté lors de la réunion de lancement, sans que cela ne compromette la capacité du MO à assurer ses contrôles.

Les unités des paramètres décrits, la référence altimétrique pour les hauteurs, et le système horaire du relevé devront être spécifiés dans une en tête. Si une de ces références n'est pas connue, cela devra être indiqué par un acronyme (Inc par exemple).

5.2.3 – Conditions d'exécution de la prestation

Les fichiers de données à numériser seront les fichiers issus de la phase de scannage, en ne traitant que ceux dont la numérisation est demandée.

Le titulaire signalera toute anomalie et tout fichier image qu'il considérera comme impropre à la numérisation avec le motif précis du rejet, dans le cadre de la vérification préalable définie à l'article 5.1.4 . Il appartiendra à la MO de décider si le document concerné doit être retiré du lot ou non.

5.2.4 – Liste des documents à numériser

Le fichier d'inventaire priorisé déjà fourni par la MO dans la phase de scannage servira de base. On exclura les fichiers dont la numérisation n'est pas demandée (information figurant dans l'inventaire). On en trouvera une description à l'annexe 2.

Si des documents techniques ou annexes sont fournis en plus ou en moins, l'inventaire devra être amendé en ce sens. Cet inventaire numérique servira pour le suivi et la traçabilité de l'action de scannage et les amendements éventuels seront complétés au fur et à mesure de l'avancement de la tâche.

5.2.5 – Nomenclature

Les règles de la nomenclature devront s'appliquer successivement :

- A l'arborescence et au nommage des répertoires.
- Aux fichiers eux-mêmes

Un exemple est proposé dans l'annexe 2. Le titulaire pourra proposer une amélioration de ces règles lors de la réunion de lancement du projet.

Les fichiers de calage et autres fichiers annexes considérés comme métadonnées devront également être soumis aux règles de nommage.

5.2.6 - Exigences techniques

Les exigences sur les documents numérisés sont les suivantes :

- Les valeurs numériques seront restituées, avec des unités et des pas de temps corrects
- Accompagnées des paramètres ayant servi au calage
- Selon la nomenclature précisée.

5.2.7 - Exécution

Le titulaire veillera à numériser les fichiers image selon les prescriptions détaillées dans l'inventaire d'accompagnement : priorité d'exécution, spécificités de la numérisation. La mise à jour des fichiers inventaire/métadonnées fait partie de la prestation de numérisation.

La numérisation se fera selon le protocole mis en place et rédigé pendant l'étape d'initialisation.

On désigne par « rapport d'exécution » l'ensemble des éléments que le titulaire aura à renseigner pour assurer le suivi et la traçabilité de son travail. Ils devront en particulier renseigner :

- les données de calage nécessaires à la numérisation
- les difficultés rencontrées le cas échéant
- pour chaque image traitée, inclure les commentaires nécessaires (par exemple lorsque l'image aura nécessité un post-traitement).
- l'état général d'avancement des tâches sous forme de tableaux

Les rapports d'exécution devront donc être accessibles en ligne par la MO, et remis à jour au moins au pas de temps mensuel. A la fin de la prestation, ils devront être inclus dans le dossier de livrables.

Le titulaire s'engage à fournir des livrables contrôlés. Le contrôle consistera essentiellement à comparer la fidélité de la série numérisée avec la donnée originale (fichier image). La réalisation du contrôle devra être consignée dans le rapport d'exécution.

5.2.8 – Procédure de vérification finale de la numérisation

Les opérations de vérification s'attacheront à comparer les documents numérisés avec les originaux et à valider la conformité des livrables avec les exigences du CCTP (voir 5.2.6).

Le délai maximal dont dispose la MO pour procéder aux opérations de vérification sera défini dans chaque bon de commande. La méthode utilisée par la MO est décrite par la norme NF ISO 2859-1. On procédera par double contrôle par échantillon. Le NQA (Niveau de Qualité Acceptable) est fixé à 2,5% (ce niveau pourra être modulé dans chaque bon de commande selon l'état du gisement). A partir d'un nombre d'anomalies constaté supérieur au seuil déduit du volume de l'échantillon, la prestation sera rejetée. Le titulaire est tenu de refaire la prestation à ses frais dans un délai de 1 mois.

Un procès-verbal sera établi par la MO à l'issue des opérations de vérification.

6 – Réunions et livrables

6.1 – Réunions

Dans l'étape d'initialisation du projet, sont prévues :

- une réunion de présentation avec la MO dans les locaux du MO ;
- la formation dispensée par le CEREMA à Toulouse
- une réunion de bilan des tests de scannage et numérisation

Dans les étapes de sauvegarde de chaque gisement de données :

- une réunion de lancement de la prestation pour présenter le gisement et les travaux à entreprendre, dans les locaux du MO ou en visio-conférence.
- une réunion intermédiaire de suivi
- la réunion finale de restitution.

6.2 – Livrables

Chaque étape fait l'objet d'un ou plusieurs livrables détaillés ci-dessous.

Dans l'étape d'initialisation du projet :

- Les documents scannés, ainsi que le compte-rendu et procès-verbal du test de scannage rédigés par le titulaire.
- Les documents numérisés, ainsi que le compte-rendu et procès-verbal du test de numérisation rédigés par le titulaire.

(note: de nouveaux comptes-rendus et procès-verbaux seront rédigés en cas d'échec des tests).

Dans les étapes de sauvegarde, pour le traitement de chaque gisement de données :

- Une synthèse mensuelle des rapports d'exécution du scannage
- Une synthèse mensuelle des rapports d'exécution de la numérisation
- Les inventaires le cas échéant
- Les fichiers intermédiaires de numérisation (fichiers de calage et fichiers bruts)
- Un rapport final de présentation
- une mise à disposition des documents scannés et numérisés sur un espace de partage sécurisé.

Annexe 1 : acronymes et définitions

Acronymes

CCAP : Cahier des Clauses Administratives Particulières

CCTP : Cahier des Clauses Techniques Particulières

DREAL : Direction Régionale de l'Environnement de l'Aménagement et du Logement

CEREMA : Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement

NUNIEAU : Logiciel gratuit et libre d'utilisation développé par le CEREMA et visant à numériser des graphiques (pluviogrammes ou hydrogrammes, par exemple).

Définitions

Pluviogramme

Courbe d'enregistrement d'un pluviographe. Elle représente l'évolution de la hauteur de précipitations en fonction du temps.

Hydrogramme

Courbe d'enregistrement d'un limnigraphe. Elle représente l'évolution de la hauteur d'eau en fonction du temps.

Registre de hauteurs d'eau, de pression, ou de vent

Recueil contenant des relevés de hauteur d'eau, de pression, et de vent (vitesse, direction, force) sous forme de tables. Ces recueils se trouvent sous forme reliée, ou sous la forme de simples feuillets.

Métadonnées

On définit les métadonnées comme des informations complémentaires aux données. Ce sont des informations relatives au contenu, à la qualité, aux sources, et toutes autres caractéristiques utiles à la qualification et l'exploitation des données qu'elles concernent. A titre d'exemple, il peut s'agir des courbes de tarage (cas des hydrogrammes) ou des fichiers décrivant les caractéristiques géographiques).

Scanner

Produire une image numérique d'un document papier.

Numériser

Produire une série de données/valeurs numériques (tableau numérique, fichier texte,...) à partir du contenu d'une image numérique (tableaux, courbes, ..).

Annexe 2: structure de fichiers d'entrée et de sortie, nomenclature

1 – Fichier d'inventaire priorisé

Ces fichiers, réalisés par la MO, seront remis au titulaire après l'émission du bon de commande correspondant. Ils devront contenir l'ensemble des informations et instructions nécessaires à la réalisation des phases de scannage et/ou de numérisation.

Chaque document papier doit ainsi pouvoir être associé à un ensemble de paramètres nécessaire à sa description et aux tâches futures. A titre d'exemple :

Modalités prises par les paramètres

n°priorité	n°interne issu de la priorisation.
Identifiant document	un identifiant unique pour le document à scanner (TPN_T_19800125_19800130)
identifiant station	n° unique permettant d'identifier le poste de mesure (10 caractères).
Station nom	Nom de la station.
Station abbréviation	Nom de la station abrégé sur 3 caractères
Type	deux modalités peuvent être prises: graphique/tableau
Format	4 modalités: inférieur ou égal à A3 / supérieur à A3 et relié/non relié
Paramètre	modalités possibles: pluie (P), hauteur cours d'eau (H), jaugeages (J)
Date début	début de la mesure
Date fin	fin de la mesure
Groupe métadonnées	Référence à un fichier de métadonnées associé (voir paragraphe suivant)
Commentaires	tout ce qui semble utile pour les phases de scannage et/ou de numérisation
Indicateur numérisation	Indique si le document doit également être numérisé.

2 – Fichier de métadonnées

Comme décrit précédemment, et conformément à la présentation générale du marché, le fichier de métadonnées contient des informations utiles à l'analyse des données mais pas nécessairement contenues dans les documents eux-mêmes. Ces informations sont globales à des groupes de documents, Il n'est pas nécessaire de les répéter in extenso pour chaque document papier. En revanche, un lien doit pouvoir être reconstitué in fine. Il peut s'agir par exemple :

- Des coordonnées des stations et son altitude
- Des unités physiques de la mesure (mm ou 1/10 de mm par exemple)
- Les pas de temps concernés (cas des tableaux)
- Des remarques génériques liées au fonctionnement des observations
- Des lacunes recensées.
- Le nombre de feuilles pour chaque format.

3 – Nomenclature

Ce qui suit constitue la structure générale du nommage des divers éléments. Des amendements et suggestions de la part du titulaire sont possibles et seront examinés lors de la réunion de démarrage du projet.

3.1 – Scannage

Nommage des répertoires

Chaque répertoire livré sera désigné par :

- Un identificateur du service demandeur, sur 2 ou 3 lettres, MO pour le Service de Prévision des Crues Méditerranée Ouest, GTL pour le Service de Prévision des Crues Garonne Tarn Lot
- Un identificateur de la phase en cours (S pour scannage ; N pour numérisation)
- une identification du paramètre traité : « P » pour pluviométrie, « H » pour hauteur d'eau, « J » pour jaugeages.

Les groupes de caractère seront séparés par des « _ » (tirets bas) de façon à produire une chaîne de caractère continue.

Par exemple, « *GTL_S_P* » pour un gisement à scanner contenant des données pluviométriques du SPC GTL.

Nommage des sous-répertoires

On ajoutera à la chaîne précédente une identification du type de document, « G » pour graphique, « T » pour tableau, ainsi que le code à 10 caractères de la station et son nom abrégé. Par exemple, « *GTL_S_P_G_1234567890_TPN* » par exemple

Nommage des fichiers de données

De façon générale, chaque fichier sera nommé à partir de :

- l'identificateur de la station de mesure, sur 10 caractères. En cas d'un identificateur comportant moins de chiffres significatifs, on ajoutera des « 0 » en début de chaîne de caractères pour compléter.
- Un identificateur de date de début d'observation, du type « AAAAMMJJ ».
- Un identificateur de date de fin d'observation du même type.

A titre d'exemple : « *1234567890_19670105_19670512* ».

Si un fichier devait être traité à nouveau pour des raisons diverses, sur la même période, on ajoutera un n° d'ordre. Par exemple, sur le cas précédent : « *34512001_19670105_19670512_2* ».

NB Pour le scan des registres mensuels, afin de rendre le nom du fichier plus lisible, on pourra ne pas faire figurer ni le jour de début ni la date de fin. Exemple : *12134567890_198001* pour le scan du registre mensuel de la station 1234567890 du mois de 01 de l'année 1980.

3.2 – Numérisation

On appliquera la même logique générale au nommage des fichiers.

Annexe 3: exemples de documents à scanner

Ce qui suit vise à fournir une première description des principaux types de données que le titulaire aura à traiter. La liste n'est pas exhaustive. Plusieurs variantes en termes de contenu ou de format physique existent dans les archives à traiter.

1 – Pluviogramme

Le pluviogramme est une bande papier de format variable où les épisodes pluvieux sont représentés par des lignes obliques plus ou moins verticales selon l'intensité de la pluie. Les épisodes secs sont formalisés par un trait horizontal. Un exemple est fourni ci-dessous. Ils proviennent d'un appareil enregistreur appelé pluviographe, constitué d'un tambour sur lequel est fixée une feuille de papier imprimée. Un stylet à encre est activé dès lors qu'une pluie est observée.

L'axe horizontal va représenter l'évolution temporelle, l'axe vertical représente l'intensité de la pluie. Ces deux valeurs peuvent être différentes selon le type d'appareil utilisé. Typiquement, un pluviogramme peut représenter une semaine d'observation, mais parfois beaucoup plus.



Figure 2 : exemple de pluviogramme

De façon générale, les pluviogrammes sont très souvent accompagnés de mentions manuscrites comme dans la figure. Souvent, il s'agit de cumuls de pluie sur une période ou un épisode donnés, reportés par l'observateur. D'autres mentions sont souvent reportées, comme des commentaires sur d'éventuels problèmes de mesure, ou bien des références liées au poste de mesure.

D'autres données, externes au document lui-même, sont d'une grande importance, comme la position géographique du poste, le type d'appareil ou diverses remarques émanant du service. Elles forment ce qu'on appelle les « métadonnées d'exploitation ». Leur forme et leur contenu diffèrent selon le service. Un lien doit être conservé entre les fichiers numérisés et ces métadonnées.

Parmi les problèmes constituant une gêne pour la numérisation figurent :

- Un effacement relatif de l'encre, lié au vieillissement.
- des « retournements » liés à l'atteinte par le stylet de la partie haute ou basse de la feuille (comme sur la figure), pouvant être multiples en cas de forte pluie.
- Le « cerisage » : Certaines données ont pu faire l'objet de traitement (cerisage), qui consiste à apposer une annotation sur les points caractéristiques de la courbe. Ces derniers sont parfois délicats à gérer dans une phase de reconnaissance de couleur pour le logiciel NUNIEAU.
- Des taches provenant de diverses origines.

Le logiciel Nunieau a pour objet de prendre en charge ces problèmes. Néanmoins, une analyse des cas particuliers et un dialogue avec la MO peuvent se révéler nécessaires. La session de formation

initiale et la documentation associée aborderont ces questions et les façons de les résoudre.

2 – Liminigramme

Un liminigramme est basé sur un principe assez semblable, mais le paramètre physique observé est une hauteur dans un cours d'eau. Un exemple de limnigraphe est fourni ci-dessous.

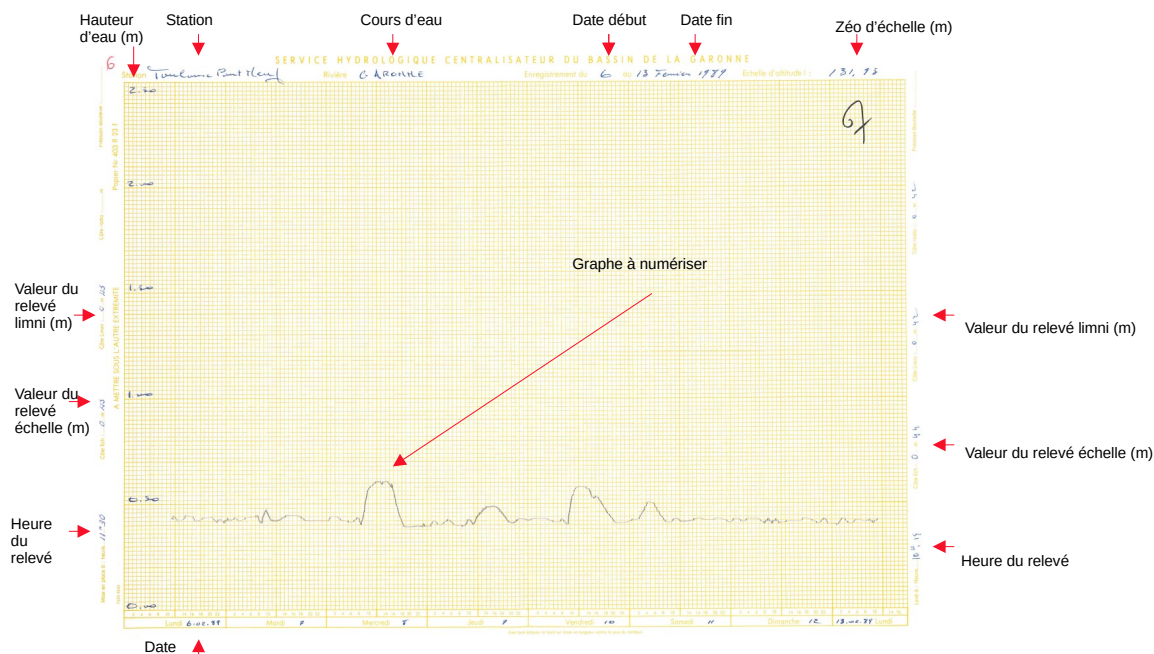


Figure 3 : exemple de liminigramme

De façon générale, les limnigrammes présentent une plus grande diversité que les pluviogrammes, notamment dans l'extension de l'axe horizontal, puisque les documents peuvent représenter de plusieurs jours à plusieurs semaines.

Généralement, pour les limnigrammes présents à a DREAL Occitanie, ils contiennent une semaine de données. Les rouleaux supports des limnigraphe étaient en effet en général changés toutes les semaines.

3 – Tableau

Les tableaux de données représentent généralement un cumul de pluie ou une hauteur d'eau moyennées sur une certaine durée (heure, jour, mois). A titre indicatif, le format le plus répandu concernant la hauteur d'eau est détaillé ci-dessous.

Année

Mois

Jour

Hauteurs d'eau (en m) à 06h00 du matin, 12h00 et 18h00 le 09/06/1900

Hauteurs d'eau en crue, reportées de façon horaire

La valeur entourée (3,55 m) correspond à la hauteur d'eau du 05/06/1900 à 02h00 du matin

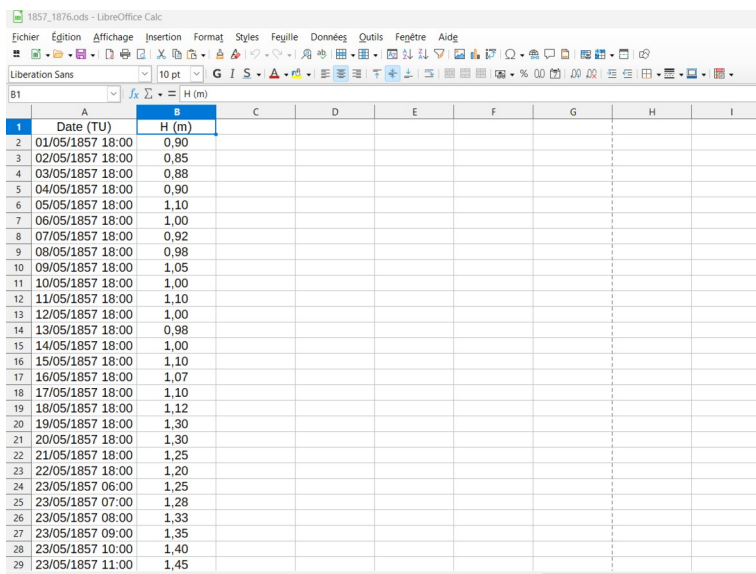
OBSERVATIONS COLLECTIVES			OBSERVATIONS DES CRUES												RENSEIGNEMENTS DIVERS
HAUTEURS D'EAU	HAUTEURS D'EAU		HAUTEURS D'EAU												DIVERS
DATE	1 h.	2 h.	1 h.	2 h.	3 h.	4 h.	5 h.	6 h.	7 h.	8 h.	9 h.	10 h.	11 h.	12 h.	
1	0.00	0.00													Notes: Hauteur d'eau à 06h00 du matin, 12h00 et 18h00 le 09/06/1900 Hauteurs d'eau en crue, reportées de façon horaire La valeur entourée (3,55 m) correspond à la hauteur d'eau du 05/06/1900 à 02h00 du matin
2	0.00	0.00													
3	0.00	0.00													
4	0.00	0.00													
5	0.00	0.00													
6	0.00	0.00													
7	0.00	0.00													
8	0.00	0.00													
9	0.00	0.00													
10	0.00	0.00													
11	0.00	0.00													
12	0.00	0.00													
13	0.00	0.00													
14	0.00	0.00													
15	0.00	0.00													
16	0.00	0.00													
17	0.00	0.00													
18	0.00	0.00													
19	0.00	0.00													
20	0.00	0.00													
21	0.00	0.00													
22	0.00	0.00													
23	0.00	0.00													
24	0.00	0.00													
25	0.00	0.00													
26	0.00	0.00													
27	0.00	0.00													
28	0.00	0.00													
29	0.00	0.00													
30	0.00	0.00													
31	0.00	0.00													

Figure 4 : exemple de tableau

Lors de l'étape de numérisation, toutes les valeurs présentes dans les tableaux devront être renseignées, avec une date et une heure devant chaque valeur. Suivant les jours, le prestataire aura donc 1, 3 ou jusqu' à 24 valeurs à renseigner.

A titre d'exemple, la numérisation de l'ensemble des relevés de la station de Toulouse Pont-Neuf, entre 05/1857 et 01/1998 a nécessité la saisie de 78 000 valeurs – (et dates/heures associées) environ.

Annexe 4: Fichier de rendu pour la numérisation des tableaux



	A	B	C	D	E	F	G	H	I
1	Date (TU)	H (m)							
2	01/05/1857 18:00	0.90							
3	02/05/1857 18:00	0.85							
4	03/05/1857 18:00	0.88							
5	04/05/1857 18:00	0.90							
6	05/05/1857 18:00	1.10							
7	06/05/1857 18:00	1.00							
8	07/05/1857 18:00	0.92							
9	08/05/1857 18:00	0.98							
10	09/05/1857 18:00	1.05							
11	10/05/1857 18:00	1.00							
12	11/05/1857 18:00	1.10							
13	12/05/1857 18:00	1.00							
14	13/05/1857 18:00	0.98							
15	14/05/1857 18:00	1.00							
16	15/05/1857 18:00	1.10							
17	16/05/1857 18:00	1.07							
18	17/05/1857 18:00	1.10							
19	18/05/1857 18:00	1.12							
20	19/05/1857 18:00	1.30							
21	20/05/1857 18:00	1.30							
22	21/05/1857 18:00	1.25							
23	22/05/1857 18:00	1.20							
24	23/05/1857 06:00	1.25							
25	23/05/1857 07:00	1.28							
26	23/05/1857 08:00	1.33							
27	23/05/1857 09:00	1.35							
28	23/05/1857 10:00	1.40							
29	23/05/1857 11:00	1.45							

Figure 5: Fichier fourni par le maître d'ouvrage pour la numérisation des tableaux : page de saisie

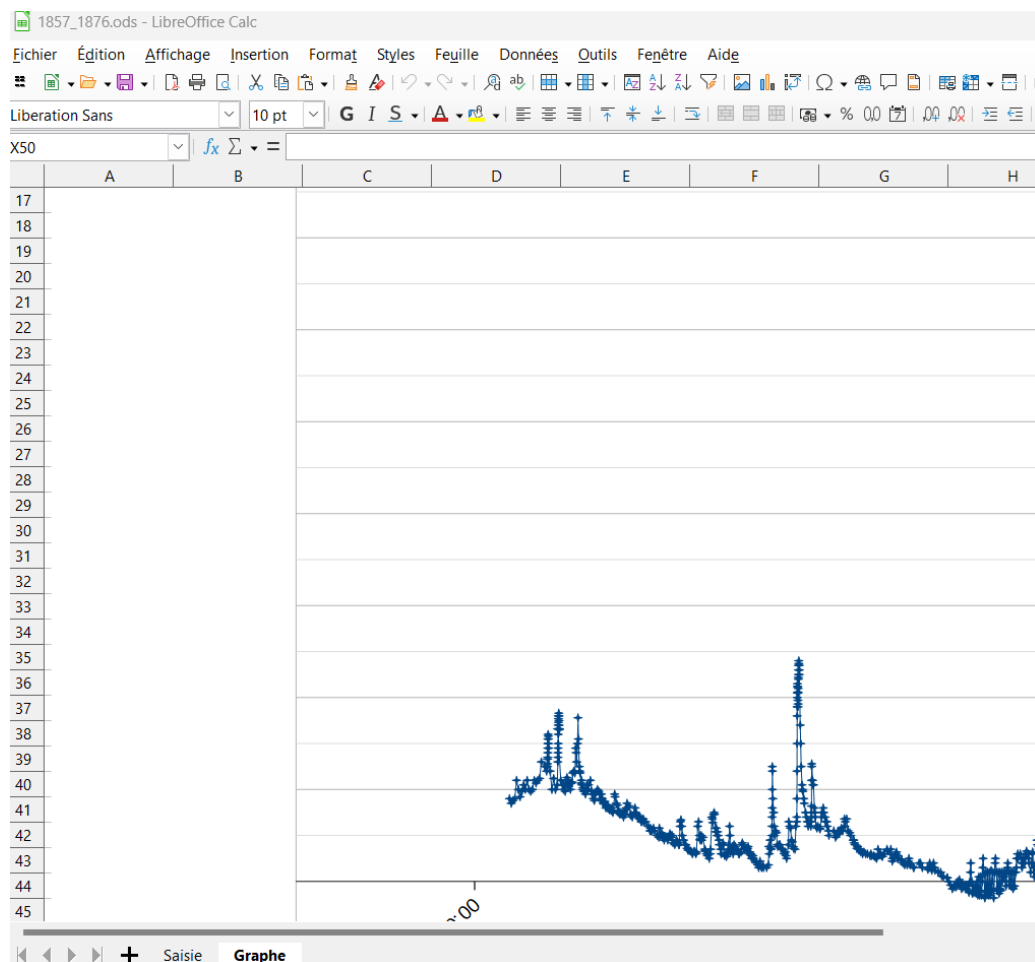


Figure 6: Fichier fourni par le maître d'ouvrage pour la numérisation des tableaux : page de contrôle