

NAADES

Données sur la qualité des eaux de surface

**Guide d'utilisation de l'outil
de contrôle des fichiers
QUESU CSV**

Statut : Validé
Diffusion :
Version : V2.2.0
Référence :

Approbation

Nom	Titre / Service	Lecteur	Valideur	Date	Version
Adrian CHARTIN	Chef de projet		X	10/09/2018	2.0.0
Lucie CLAVEL	Cheffe de projet		X	01/07/2019	2.1.3
Grégoire ETOT	Directeur de projet		X	01/04/2020	2.2.0

Documents référencés

N°	Version	Date	Auteur	Description
01	3.1	05/05/2017	SANDRE	Scénario d'échanges de données (TEXT/CSV) Qualité des eaux superficielles continentales (QUESU CSV PHY 3.1)
02	03	25/02/2019	AFB	CR du GT Producteur du 25/02/2019
03		25/02/2019	AFB	Spécifications QUESU CSV PHY adaptées à l'échange Banque de référence => NAIADES validées par le GT Producteur du 25/02/2019

Révision documentaire

Version	Date	Rédigé par	Vérifié par :	Validé par : (Fonction / Nom)	Description
1.0	06/06/2017	A. ANDRADE	L. CLAVEL	L. CLAVEL	Création du document
1.1	24/07/2017	A. ANDRADE	L. CLAVEL	L. CLAVEL	Complément à l'installation de l'outil
1.2	04/12/2017	A. ANDRADE	L. CLAVEL	L. CLAVEL	Installation sous debian 9.2
2	07/09/2018	A. ANDRADE	A. CHARTIN	A. CHARTIN	Prise en compte de l'exécution de l'outil par NAIADES
2.1.3	01/07/2019	A. ANDRADE	L. CLAVEL	L. CLAVEL	Prise en compte des spécifications [R03]

Version	Date	Rédigé par	Vérifié par :	Validé par : (Fonction / Nom)	Description
2.2.0	01/04/2020	A. ANDRADE			Création de l'OFB Support Python 3.7 et supérieur Migration API Référentiel Sandre

SOMMAIRE

1. OBJET DU DOCUMENT.....	5
2. DESCRIPTION DES CONTROLES	6
3. INSTALLATION DE L'OUTIL	7
3.1 Prérequis.....	7
3.2 Installation.....	7
3.3 Description des fichiers installés	8
4. UTILISATION DE L'OUTIL	9
4.1 Contrôler des fichiers QUESU CSV.....	9
4.2 Interpréter l'arborescence des fichiers résultant d'un contrôle	10
4.3 Interpréter le contenu d'un fichier de non-conformités	11
ANNEXE – SERVICE WEB UTILISES POUR LE CONTROLE	14

1. OBJET DU DOCUMENT

L'application web [NAIADES](#) met à disposition du public les données disponibles relatives à la qualité des eaux de surface continentales. Parmi ces données, le sous-ensemble des données de physicochimie peut être mis à jour via l'import de fichiers conformes à la [spécification SANDRE QUESU CSV PHY 3.1](#).

Pour optimiser les échanges avec les producteurs de données, l'équipe projet OFB / BRGM met à disposition un outil de contrôle de la syntaxe et de la sémantique des fichiers QUESU CSV. Ce document décrit l'installation et l'utilisation de cet outil. Il présente également les contrôles exécutés.

2. DESCRIPTION DES CONTROLES

L'outil parcourt le contenu d'un répertoire désigné par l'utilisateur et exécute les contrôles suivants sur chaque fichier analysé :

1. Validation de la règle de nommage d'un fichier QUESU CSV :
 - *QUESU_CSV_ANA_*.csv* pour un fichier d'analyses de physicochimie ;
 - *QUESU_CSV_CEP_*.csv* pour un fichier de conditions environnementales de physicochimie.

Les fichiers QUESU CSV compressés ne sont pas pris en charge. Les fichiers mal nommés sont ignorés pour la suite des opérations.

2. Vérification de l'encodage du fichier QUESU CSV :
 - Encodage *utf8* des données ;
 - En-tête limité à la 1^{ère} ligne des intitulés de colonnes ;
 - Absence de caractères de retour à la ligne dans les données.

Pour plus d'efficacité, l'outil prend en charge les traitements correctifs :

- Ré-encodage des fichiers en *UTF8* ;
- Suppression de la 2ème ligne d'en-tête lorsqu'elle est détectée. Compte tenu de la difficulté de cette détection, il est fortement recommandé de limiter l'en-tête à la ligne des intitulés de colonnes ;
- Suppression de tout caractère de retour à la ligne présent dans la donnée.

3. Vérification de la syntaxe QUESU CSV

Cette vérification se base sur les spécifications validées par le GT Producteur (cf. [REF03]).

4. Validation des références SANDRE

Cette opération consiste à vérifier l'existence d'une référence diffusée par les services web du SANDRE (cf. liste en fin de document)

En cas d'échec du téléchargement des référentiels, les fichiers xml existants sur le poste de travail sont utilisés. Pour plus d'efficacité, l'outil limite l'interrogation des services web à une fois par jour.

En cas d'échec d'une des étapes, les opérations suivantes ne sont pas exécutées. Une même donnée peut générer plusieurs non-conformités (exemple : format invalide pour deux données élémentaires d'une même analyse) Les non-conformités sont enregistrées dans des fichiers au format CSV (cf. description plus bas)

3. INSTALLATION DE L'OUTIL

3.1 Prérequis

L'utilisation de l'outil de contrôle requiert une connexion internet permettant l'interrogation des services web du SANDRE.

3.2 Installation

Windows :

1. Installer l'interpréteur python version 3.7 ou supérieure (installateurs disponibles sous <https://www.python.org/downloads/>)
2. Créer un répertoire pour l'installation de l'outil de contrôle (exemple : `C:\naiades\quesu_csv_validator_2_2_0`)
3. Décompresser le contenu de l'archive `quesu_csv_validator_2_2_0.zip` dans le répertoire d'installation.
4. Installer les modules requis pour l'exécution de l'outil contrôle
 - i. Ouvrir une fenêtre console
 - ii. Naviguer vers le répertoire d'installation

```
cd C:\naiades\quesu_csv_validator_2_2_0
```

- iii. Installer les modules

```
pip install -r requirements.txt
```

5. Patienter durant l'installation des modules
6. Fermer la fenêtre console

debian 9.2 :

1. Vérifier l'installation de l'interpréteur python en version 3.5 ou supérieure
 - i. Ouvrir une fenêtre console
 - ii. Interroger la version de l'interpréteur

```
python3 -v
```

2. Créer un répertoire pour l'installation de l'outil de contrôle

```
mkdir /var/naiades/quesu_csv_validator_2_2_0
```

3. Décompresser le contenu de l'archive *quesu_csv_validator_2_2_0.zip* dans le répertoire d'installation.

```
cd unzip quesu_csv_validator_2_2_0.zip
```

4. Installer les modules requis pour l'exécution de l'outil contrôle

```
pip3 install -r requirements.txt
```

5. Patienter durant l'installation des modules
6. Fermer la fenêtre console

3.3 Description des fichiers installés

Une fois l'installation réalisée, les fichiers suivants sont disponibles :

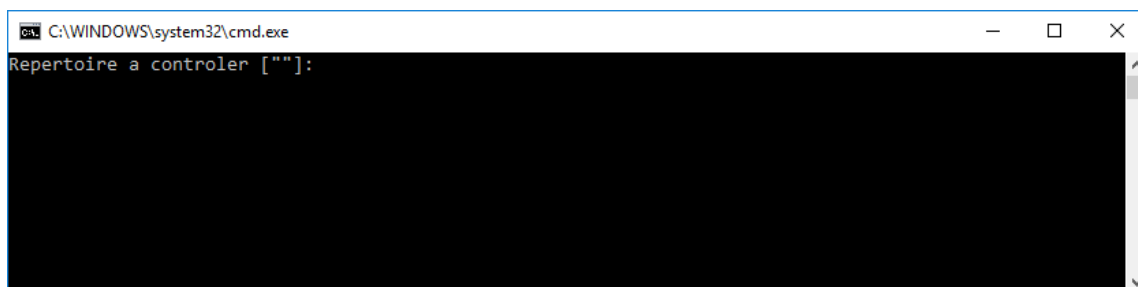
Fichier	Description
quesu_csv_validator.bat	Script de commande facilitant l'exécution de l'outil de contrôle en environnement Windows
quesu_csv_validator.py	Module principal de l'application python
quesu_csv_file.py	Module de parcourt des fichiers QUESU CSV d'un répertoire
quesu_csv_line.py	Module de contrôle d'une ligne de données QUESU CSV
quesu_csv_reference.py	Module de téléchargement des référentiels SANDRE utilisés pour le contrôle des fichiers QUESU CSV
quesu_csv_error.py	Module d'enregistrement des non conformités détectées lors du contrôle
quesu_csv_type.py	Module de description des types de données QUESU supportés
Dockerfile	Script de configuration nécessaire pour une exécution de l'outil de contrôle par NAIADES
naiades.config	Script de configuration nécessaire pour une exécution de l'outil de contrôle par NAIADES
webservice.py	Module nécessaire pour une exécution de l'outil de contrôle par NAIADES
requirements.txt	Listing des modules nécessaires pour l'exécution de l'outil de contrôle
quesu_csv_validator_2_2_0.pdf	Guide d'utilisation de l'outil de contrôle

4. UTILISATION DE L'OUTIL

4.1 Contrôler des fichiers QUESU CSV

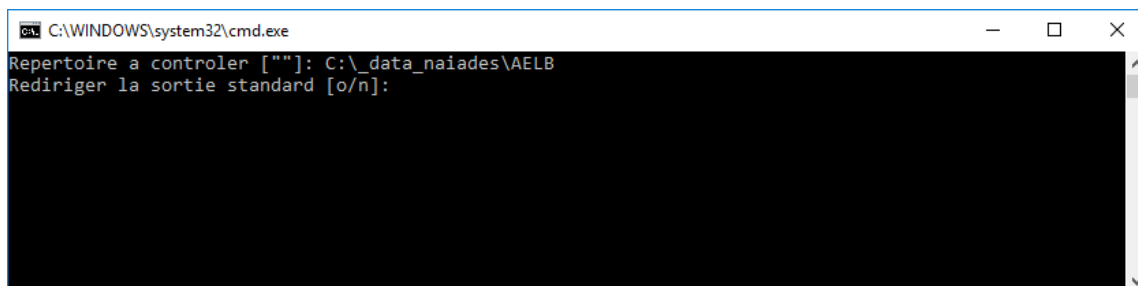
Windows :

1. Exécuter le script de commande DOS *quesu_csv_validator.bat*. Une fenêtre s'affiche invitant l'utilisateur à renseigner le chemin d'accès au répertoire à contrôler :



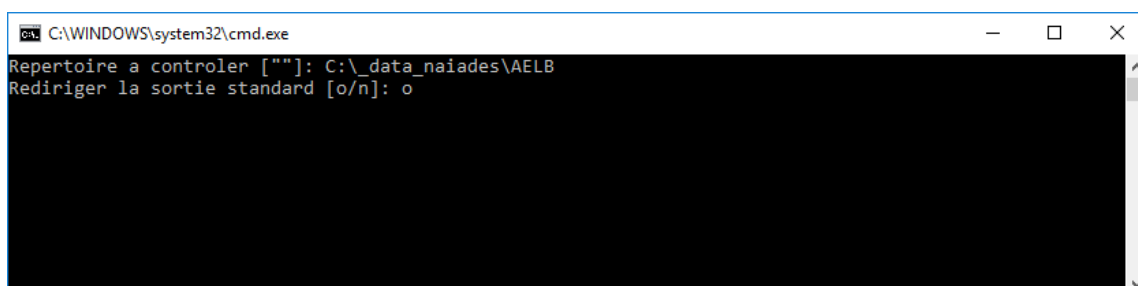
```
C:\WINDOWS\system32\cmd.exe
Repertoire a controler [""]:
```

2. Renseigner le chemin d'accès puis taper la touche *Entrée*. L'utilisateur est alors invité à rediriger la sortie standard du script python :



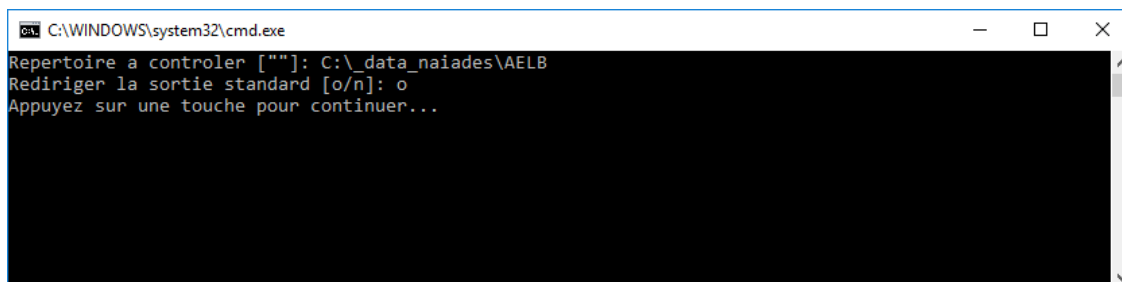
```
C:\WINDOWS\system32\cmd.exe
Repertoire a controler [""]: C:\_data_naiades\AELB
Rediriger la sortie standard [o/n]:
```

3. Renseigner *o* pour journaliser les traitements puis taper la touche *Entrée*. Le traitement s'exécute en tâche de fond. En fonction du volume de fichiers, la durée du contrôle peut varier entre 25 min et 2j.



```
C:\WINDOWS\system32\cmd.exe
Repertoire a controler [""]: C:\_data_naiades\AELB
Rediriger la sortie standard [o/n]: o
```

- Une fois le contrôle achevé, l'outil invite l'utilisateur à taper la touche *Entrée*. pour fermer la console d'exécution :



```
C:\WINDOWS\system32\cmd.exe
Repertoire a controler [""]: C:\_data_naiades\AELB
Rediriger la sortie standard [o/n]: o
Appuyez sur une touche pour continuer...
```

- Taper *Entrée* pour fermer la console.

Debian 9.2 :

- Ouvrir une fenêtre terminal
- Naviguer vers le répertoire d'installation de l'outil de contrôle

```
cd /var/naiades/quesu_csv_validator_2_2_0
```

- Définir l'encodage *UTF-8* pour l'interpréteur python

```
export PYTHONIOENCODING=utf-8
```

- Exécuter l'interpréteur python en spécifiant en paramètres le module principal de l'application et le répertoire des fichiers QUESU CSV à contrôler :

```
python3 quesu_csv_validator.py -d [chemin d'accès au répertoire des fichiers QUESU CSV] 1>[chemin d'accès à un fichier de log]
```

4.2 Interpréter l'arborescence des fichiers résultant d'un contrôle

A l'issue d'un contrôle, le répertoire des fichiers QUESU CSV analysé contient trois sous répertoires supplémentaires :

Répertoire	Description
quesu_validation	Répertoire regroupant les fichiers QUESU CSV contrôlés. En fonction des données d'entrée, ces fichiers diffèrent des fichiers d'origine par l'encodage, l'en-tête de données QUESU et/ou l'absence de retour à la ligne dans les données.

Répertoire	Description
quesu_reference	Répertoire regroupant les fichiers XML et gzip des référentiels SANDRE utilisés pour le contrôle des fichiers QUESU CSV PHY.
quesu_error	Répertoire regroupant les fichiers CSV de non conformités détectées lors du contrôle.

Recommandation : pour faciliter l'exploitation des résultats, il convient d'effacer les répertoires *quesu_validation* et *quesu_error* entre deux exécutions de l'outil.

Si, comme recommandé, les sorties standards du script ont été redirigées vers des fichiers, le répertoire des fichiers QUESU CSV contient deux autres fichiers :

Fichier	Description
quesu_csv_validation.err	Fichier d'enregistrement de la sortie pour les messages d'erreur.
quesu_csv_validation.log	Fichier d'enregistrement de la sortie pour les autres messages.

4.3 Interpréter le contenu d'un fichier de non-conformités

La description d'une non-conformité est enregistrée dans un fichier du répertoire *quesu_error* suivant le format CSV suivant :

Fichier;Gravite;Erreur;Ligne;Colonne;Champ;Valeur;DonneesQUESU

Le tableau suivant décrit chaque champ de données :

Champ	Description
Fichier	Chemin d'accès au fichier QUESU CSV contrôlé
Gravite	Criticité de la non-conformité
Erreur	Libellé de la non-conformité
Ligne	N° de la ligne contenant la non-conformité dans le fichier QUESU CSV contrôlé
Colonne	N° de la colonne contenant la non-conformité dans le fichier QUESU CSV
Champ	Libellé du champ QUESU CSV concerné par la non-conformité
Valeur	Valeur du champ QUESU CSV concerné par la non-conformité
DonneesQUESU	Données QUESU de l'observation (analyse ou condition environnementale) concernée par la non-conformité

L'outil distingue trois niveaux de gravité pour les non-conformités :

- **BLOQUANTE** : ce type de non-conformité bloque le processus d'alimentation. Elle est rencontrée dans les cas de figure suivants :

Origine de la non-conformité	Libellé de l'erreur dans le fichier de non-conformités
Nombre de champs incorrect pour l'en-tête	EN-TETE INCORRECT
Nombre de champs incorrect pour les données (en particulier, en présence d'un « ; » dans les commentaires)	NOMBRE DE CHAMPS INCORRECT
Donnée élémentaire obligatoire absente (exemple : code de la station de mesure)	DONNEE ABSENTE
Donnée obligatoire mal formatée (exemple : code de la station de mesure de longueur supérieure à 8 caractères)	FORMAT INCORRECT
Donnée obligatoire correspondant à une référence Sandre non valide	REFERENCE INVALIDE

Le traitement de ce type de non-conformité impose les actions suivantes du producteur des données :

- 1) Corriger les données csv et/ou les données du SI de production
- 2) Extraire, si nécessaire, un nouveau lot de données csv
- 3) Relancer le contrôle des données

- **NON BLOQUANTE** : ce type de non-conformité ne bloque pas le processus d'alimentation. Elle est rencontrée dans les cas de figure suivants :

Origine de la non-conformité	Libellé de l'erreur dans le fichier de non-conformités
Fichier non encodé en « utf8 »	ENCODAGE INCORRECT
En-tête sur deux lignes (intitulés des champs sur la 1 ^{ère} ligne, description des champs sur la 2 ^{ème} ligne)	EN-TETE INCORRECT
Donnée facultative mal formatée	FORMAT INCORRECT
Donnée facultative correspondant à une référence Sandre non valide	REFERENCE INVALIDE

Lorsque ce type de non-conformité est rencontré, il est recommandé au producteur d'exécuter les actions imposées pour le traitement des erreurs bloquantes.

- **INFORMATION** : ce type de non-conformité ne bloque pas le processus d'alimentation. Elle est rencontrée dans les cas de figure suivants :

Origine de la non-conformité	Libellé de l'erreur dans le fichier de non-conformités
Donnée facultative mal formatée	FORMAT INCORRECT
Donnée facultative correspondant à une référence Sandre non valide	REFERENCE INVALIDE

Afin maximiser la qualité des données diffusées, il est recommandé au producteur des données d'exécuter les actions imposées pour le traitement des erreurs bloquantes.

ANNEXE – SERVICE WEB UTILISES POUR LE CONTROLE

Code du référentiel	URL de téléchargement du référentiel au format XML
STQ	https://api.sandre.eaufrance.fr/referentiels/v1/stq.xml?outputSchema=SANDREv3.1&compress=true
DC	https://api.sandre.eaufrance.fr/referentiels/v1/dc.xml?outputSchema=SANDREv4&compress=true
22 (Projet Prel)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E22%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
430 (Zone Verticale Prospective)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E430%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
67 (Difficulté Prel)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E67%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
333 (Accred Prel)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E333%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
645 (Finalité Prel)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E645%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
SUP	https://api.sandre.eaufrance.fr/referentiels/v1/sup.xml?outputSchema=SANDREv2&compress=true
MET	https://api.sandre.eaufrance.fr/referentiels/v1/met.xml?outputSchema=SANDREv2&compress=true
INT	https://api.sandre.eaufrance.fr/referentiels/v1/int.xml?outputSchema=SANDREv2&compress=true
PAR	https://api.sandre.eaufrance.fr/referentiels/v1/par.xml?outputSchema=SANDREv2&compress=true

Code du référé ntiel	URL de téléchargement du référentiel au format XML
	mpress=true
FAN	https://api.sandre.eaufrance.fr/referentiels/v1/fan.xml?outputSchema=SANDREv2&compress=true
URF	https://api.sandre.eaufrance.fr/referentiels/v1/urf.xml?outputSchema=SANDREv2&compress=true
155 (Rq)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E155%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
156 (Insitu Ana)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E156%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
43 (Diffic ulteAn a)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E43%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
414 (Qual)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E414%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
446 (Statut)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E446%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E
299 (Accre Ana)	https://api.sandre.eaufrance.fr/referentiels/v1/nsa.xml?outputSchema=SANDREv3.1&compress=true&filter=%3CFilter%3E%3CIS%3E%3CField%3ECdReferentiel%3C%2FField%3E%3CValue%3E299%3C%2FValue%3E%3C%2FIS%3E%3C%2FFilter%3E