



IT LINK
Accélérateur d'innovation

Réf : GEODAIR-SFD02-001 v1.00

Edité le : 27/11/2024



TMA GEOD'AIR : Mise à jour du flux E2 vers agence européenne et data.gouv.fr

Spécifications fonctionnelles détaillées



1. Historique des versions

Version	Date	Rédacteur	Description
1.01	01/07/2020	Blandine Seznec	Version initiale
1.02	18/08/2020	Blandine Seznec	Mise à jour suite à la revue de l'INERIS.
2.00	27/08/2020	Blandine Seznec	Mise à jour suite à la V2.0 du cahier des charges
2.01	04/09/2020	Blandine Seznec	Modifications techniques
3.00	18/08/2022	Blandine Seznec	Mise à jour suite au Mantis 7612
3.01	10/11/2022	Jordan Carel	Mise à jour suite au Mantis 8062
3.03	27/01/2023	Jordan CAREL et Camille Pignon	Mise à jour suite au Mantis 8728
4.00	24/04/2024	Jordan CAREL et Olivier Vuillier	Mise à suite au Mantis 10034

Table des matières

1. Historique des versions	2
2. Introduction	4
2.1. <i>Objet du document</i>	4
3. Généralités	5
3.1.1. Contenu des datasets E2	5
3.1.2. Modalités de génération des datasets E2	5
3.1.3. Mise à disposition des datasets E2	5
3.1.4. Convention de nommage des datasets E2	5
4. Spécifications détaillées	7
4.1. <i>Génération et mise à jour des datasets E2</i>	7
4.1.1. [SP-F-0010] – Collecter les dates des datasets E2 à mettre à jour au fil de l’eau	7
4.1.2. [SP-F-0020] – Générer manuellement des datasets E2 via appel au script GenerationE2.sh	8
4.1.3. [SP-F-0030] – Générer périodiquement des datasets E2 de manière automatique	9
4.1.4. [SP-F-0040] – Générer les datasets E2 au format CSV et XML à l’aide du job generation_E2	10
4.1.5. [SP-F-0050] – Paramétrer le job Talend generation_E2	16
4.1.6. [SP-F-0060] – Générer un dataset E2 via l’API REST du module d’export CSV	16
4.2. <i>Mise à disposition des datasets E2</i>	17
4.2.1. [SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh	18
4.2.2. [SP-F-0080] – Transférer périodiquement les datasets E2 vers les serveurs cibles	19
4.2.3. [SP-F-0090] – Configurer le script PushReport.sh	19
4.2.4. [SP-F-0100] – Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr via le script pushDataGouv.sh	20
4.2.5. [SP-F-0110] – Transférer les datasets E2 sur le serveur FTP de l’Agence Européenne	21
4.2.6. [SP-F-0120] – Alerter l’administrateur en cas d’échec de transfert	21
4.2.7. [SP-F-0130] – Mettre à jour l’e-mail de rapport quotidien	21
5. Couverture des exigences	23
6. Les documents en référence	27
6.1. <i>Fourniture INERIS</i>	27
6.2. <i>Fourniture IT Link</i>	27

2. Introduction

2.1. Objet du document

Dans le cadre du rapportage des données de mesures réglementaires à la Commission Européenne, Geod'air (en application des directives 2004/107/CE et 2008/50/CE ainsi que de la décision de l'Union Européenne 2011/850/EU relative à l'échange d'information et au rapportage des données sur la qualité de l'air), transmet en temps réel à l'AEE les données automatiques de mesure des polluants réglementés sur l'ensemble du territoire. Celles-ci sont placées dans un fichier appelé E2 au format XML.

L'Agence Européenne rencontre actuellement des difficultés pour intégrer l'abondance des données fournit par l'INERIS. En effet, depuis la mise à jour pour datagouv, l'INERIS envoie pour un jour calendaire l'ensemble des données françaises dès lors qu'une modification a été apportée sur la journée considérée. L'Agence Européenne se retrouve alors, dans certains cas, à devoir réintégrer des données complètes sur plusieurs mois.

Afin de palier à ce problème, il a été décidé de procéder à une optimisation fonctionnelle du flux des E2 vers l'Europe.

Ce document consitue le dossier des spécifications fonctionnelles détaillées du processus de génération et de mise à disposition des Datasets E2 par la plateforme GEOD'AIR.

3. Généralités

3.1.1. Contenu des datasets E2

Un dataset E2 complet rapporte les moyennes horaires réglementaires (brutes et validées) de la journée de l'ensemble des points de prélèvements (SPO) des régions surveillées par le dispositif.

Un dataset E2 partiel rapporte uniquement le bloc des moyennes horaires réglementaires (brutes et validées) de la journée et dont la date d'insertion (ou de mise à jour) dans Geod'Air est postérieure à la date de génération du dernier dataset E2. Si aucun dataset E2 partiel n'a encore été généré pour la journée, il faut prendre en compte les moyennes établies depuis 00h00.

3.1.2. Modalités de génération des datasets E2

Il existe deux modalités de génération des datasets E2.

Le mode « auto » est exécutée périodiquement de manière automatique afin de rapporter les données horaires intégrées au fil de l'eau.

En mode « auto », les fichiers E2 générés pour une date donnée sont les suivants :

- Dataset complet au format XML
- Dataset complet au format CSV
- Dataset partiel au format XML : contient uniquement les moyennes horaires des SPO ayant été mises à jour depuis la dernière génération (nouvelle mesure ou modification d'une valeur). Les datasets E2 partiels ne sont transmis qu'à l'agence Européenne.

Le mode « manu » est une exécution manuelle permettant de régénérer des rapports sur une période antérieure. L'application principale identifiée pour ce second mode est d'initialiser l'espace files.data.gouv.fr.

En mode « manu », les fichiers E2 générés pour une date donnée sont les suivants :

- Dataset complet au format XML
- Dataset complet au format CSV

3.1.3. Mise à disposition des datasets E2

Tous les fichiers E2 générés sont stockés temporairement sur DATA1, jusqu'à envoi vers les serveurs cibles.

Les derniers datasets E2 générés par GEOD'AIR (mode « auto » ou « manu ») sont transférés de manière périodique.

Pour l'Agence Européenne et pour la plateforme data.gouv.fr, le transfert se fait toutes les heures.

3.1.4. Convention de nommage des datasets E2

Un dataset E2 complet est nommé de la manière suivante : FR_E2_AAAA-MM-JJ.YYY :

- AAAA : représente l'année des données présentes dans le fichier,
- MM : représente le numéro du mois des données présentes dans le fichier,

Spécifications fonctionnelles détaillées

- JJ : représente le numéro du jour dans le mois des données présentes dans le fichier,
- YYY : extension relative au format de fichier « xml » ou « csv »

Un dataset E2 partiel est nommé de la manière suivante : FR_E2_XX_T_P_AAAA-MM-JJ-hhmmss.YYY :

- AAAA : représente l'année des données présentes dans le fichier,
- MM : représente le numéro du mois des données présentes dans le fichier,
- JJ : représente le numéro du jour dans le mois des données présentes dans le fichier,
- hhmmss : représente l'heure de génération du rapport
- YYY : extension relative au format de fichier « xml »
- XX : code du réseau de mesure
- T : pour le type de la donnée « NV (non validées) » ou « V (validées) »

La lettre « P » est rajoutée entre la mention « E2 » et l'année pour indiquer qu'il s'agit d'un dataset partiel.

4. Spécifications détaillées

4.1. Génération et mise à jour des datasets E2

4.1.1. [SP-F-0010] – Collecter les dates des datasets E2 à mettre à jour au fil de l'eau

Les données automatiques sont intégrées à la base MongoDB au fil de l'eau par le job Talend « import_donnees_auto_brutes » (ou « import_donnees_auto_validees ») : job Talend d'intégration des données automatiques.

Par ailleurs, ce job est chargé de détecter si une génération de dataset E2 est nécessaire. Les dates nécessitant une génération de dataset E2 sont stockées en base de données PostgreSQL.

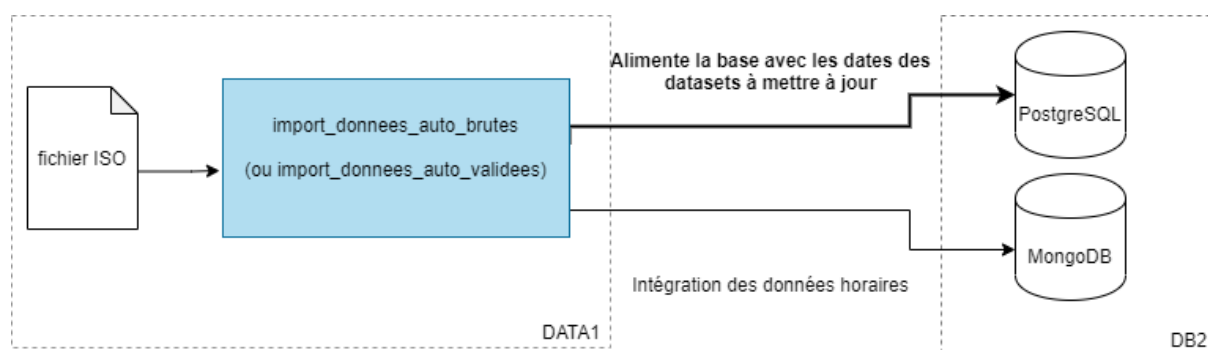


Figure 1 : Vue simplifiée du fonctionnement des ETL d'intégration des données automatiques

Désignation	Collecter les dates des datasets E2 à mettre à jour au fil de l'eau
Exigences	EX-0010 EX-0150
Accès	Fonctionnalité exécutée par les ETL « impot_donnees_auto_brute » et import_donnees_auto_validees »
Description	La collecte des dates est effectuée par les ETL d'intégration des données automatiques : import_donnees_auto_brutes et import_donnees_auto_validees (pas le job de reprise). Les dates des datasets E2 nécessitant une mise à jour sont stockées dans les base PostgreSQL dédiée <i>rapport_ue</i> et <i>rapport_historique</i> . Cf. illustration Figure 1
Entrées	Fichier ISO de données automatiques consommés par l'ETL d'intégration après dépôt sur le serveur FTP.
Paramètres	NA
Traitements	A l'issue de l'intégration des données horaires, l'ETL met à jour la table PostgreSQL dédiée au suivi des datasets E2 (cf. § « Sorties » plus bas) si une mise à jour ou génération de rapport E2 est nécessaire. Les conditions pour qu'un rapport soit généré ou mis à jour, et donc que les données soient enregistrées, sont les suivantes : <ul style="list-style-type: none"> Les moyennes horaires issues de configuration de mesures réglementaires ont été mises à jour pour cette date OU de nouvelles moyennes horaires issues de configuration de mesures réglementaires ont été intégrées pour cette date

	<p>Pour une journée donnée, trois rapports doivent être générés : un au format CSV complet, un au format xml complet et un au format xml partiel (contient uniquement les moyennes horaires des SPO ayant été mises à jour depuis la dernière génération : nouvelle mesure ou modification d'une valeur).</p>
Sorties	<p>La table <i>rapport_ue</i> est alimentée avec les champs suivants :</p> <ul style="list-style-type: none"> • Inspire_id : nom du rapport à générer (cf. 3.1.4) • Inspire_localid : date du rapport à générer au format « aaaa-MM-jj » • Inspire_namespace : vide • Inspire_version : Version incrémentée à chaque régénération. S'il s'agit de la première génération du rapport, la version est « 1 » • Statut : « En attente » • Date_creation : vide (sera rempli à la génération du rapport) • Fichier_rapportage : vide • Dataset : « E2 » • Commentaire : vide • Annee_generation : année du rapport à générer au format « AAAA » • Code_polluant : vide <p>La table <i>rapport_historique</i> est alimentée avec les champs suivants :</p> <ul style="list-style-type: none"> • Id_rapport_ue : id du dataset (table <i>rapport_ue</i>) • Action : « Génération du rapport » • Date_création : vide • Date_debut : vide • Date_fin : vide • Id_contact : 114 (Administrateur GEOD'AIR) • Nb_objets : vide

4.1.2. [SP-F-0020] – Générer manuellement des datasets E2 via appel au script GenerationE2.sh

En mode « manu », les datasets E2 sont générés via appel à un script shell *GenerationE2.sh* faisant appel aux mêmes modules que le flux automatique.

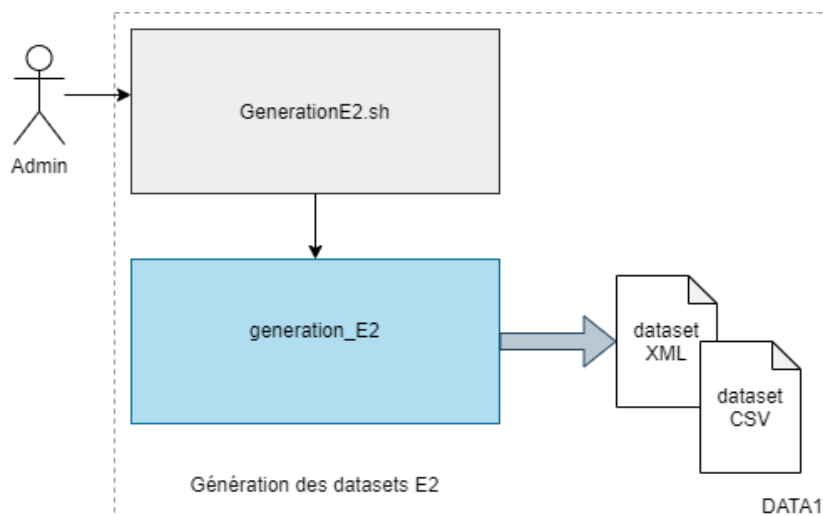


Figure 2 : Mode manuel de génération des datasets E2

En mode « manu », la liste des dates nécessitant une mise à jour des datasets E2 est renseignée par l'administrateur en paramètre du script *GenerationE2.sh*.

Désignation	Générer manuellement des datasets E2 via appel au script <i>GenerationE2.sh</i>
Exigences	EX-0030
Accès	Commande bash exécutable depuis DATA1
Description	Le script bash de génération des datasets E2 est exécutable manuellement via une commande depuis un terminal. Cf. illustration Figure 2
Entrées	NA
Paramètres	Arguments obligatoires de la commande bash : <ul style="list-style-type: none"> • Date de début de la période au format aaaa-MM-jj • Date de fin de la période au format aaaa-MM-jj Arguments optionnels : <ul style="list-style-type: none"> • --datagouv : les datasets générés seront déposés dans le répertoire d'envoi vers data.gouv.fr • --ue : les datasets générés seront déposés dans le répertoire d'envoi vers l'agence européenne
Traitements	Exécution du job de génération des datasets E2 (job Talend <i>generation_E2</i>) en mode « manu » : cf. [SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job <i>generation_E2</i> . Les paramètres d'exécution du job Talend sont les suivants : <ul style="list-style-type: none"> • Mode de génération manuel : <i>true</i> • Chemin du fichier de configuration : chemin vers le fichier de configuration défini dans la spécification [SP-F-0050] – Paramétrer le job Talend <i>generation_E2</i> • Date de début de la période : date fournie en paramètre du script • Date de fin de la période : date fournie en paramètre du script • Transfert vers data.gouv.fr : <i>true</i> si le paramètre --datagouv a été renseigné, <i>false</i> sinon • Transfert vers l'agence européenne : <i>true</i> si le paramètre --ue a été renseigné, <i>false</i> sinon
Sorties	Les logs d'exécution du scripts sont consignés dans un fichier de sortie dont le chemin est défini dans le fichier de configuration du script.

4.1.3. [SP-F-0030] – Générer périodiquement des datasets E2 de manière automatique

En mode « auto », le job Talend de génération des datasets E2 *generation_E2* est exécuté périodiquement par une tâche CRON de manière automatique (cf. Figure 3)

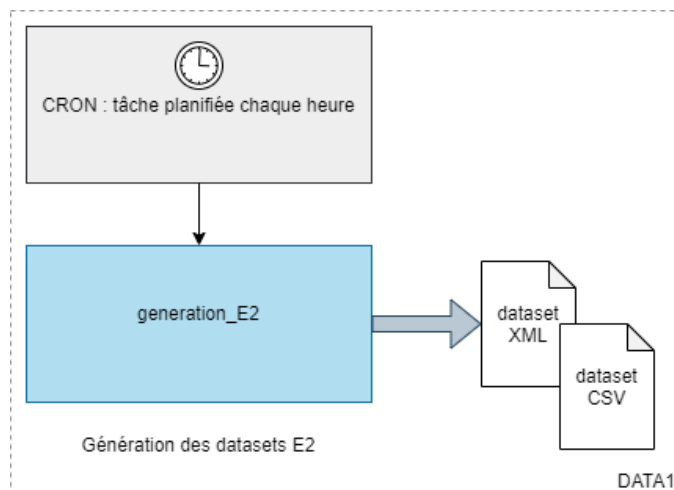


Figure 3 : Mode automatique de génération des datasets E2

La différence entre les deux modes de génération de datasets E2 réside dans la collecte des dates des datasets à générer ou à mettre à jour. En mode automatique, la liste des datasets nécessitant une mise à jour est extraite depuis la base PostgreSQL.

Désignation	Générer périodiquement des datasets E2 de manière automatique
Exigences	EX-0010
Accès	Tâche exécuté automatiquement de manière périodique
Description	Le job de génération des datasets E2 est exécuté <u>toutes les heures</u> en mode « auto » au travers d'une tâche CRON qui appelle le script Generation_E2_auto.sh. Cf. illustration Figure 3
Entrées	NA
Paramètres	NA
Traitements	Exécution du job de génération des datasets E2 (job Talend <i>generation_E2</i>) en mode « auto » : cf. [SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job <i>generation_E2</i> . Les paramètres d'exécution du job Talend sont les suivants : <ul style="list-style-type: none"> • Mode de génération manuel : <i>false</i> • Chemin du fichier de configuration : cf. [SP-F-0010] – Collecter les dates des datasets E2 à mettre à jour au fil de l'eau • Transfert vers data.gouv.fr : <i>true</i> • Transfert vers l'agence européenne : <i>true</i>. <ul style="list-style-type: none"> ○ Seuls les datasets E2 partiels sont transférés vers l'agence européenne cf. [SP-F-0110] – Transférer les datasets E2 sur le serveur FTP de l'Agence Européenne
Sorties	NA

4.1.4. [SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job *generation_E2*

Les diagrammes (Figure 4, Figure 5 et Figure 6) ci-dessous illustrent les étapes de traitement du job Talend *generation_E2*. Les acteurs et modules intervenants dans ce processus sont les suivants :

- CRON : gestionnaire de tâches périodique
- « generation_E2 » : job Talend de génération de dataset E2 au format XML (complet et partiel)
- « Module d'export CSV » : service d'export des données horaires au format CSV (complet). Ce service fait appel au module d'export CSV accessible depuis l'IHM de GEOD'AIR

Les étapes du processus sont les suivantes :

Etape 1 : Le job Talend *generation_E2* est appelé (en mode « auto », le job est appelé par une tâche CRON)

Etape 4 : Le job Talend *generation_E2* demande la génération d'un dataset E2 CSV complet via une requête HTTP au module d'export CSV

Etape 5-10 : Le module d'export CSV génère un fichier CSV complet contenant la liste des moyennes horaires issues de configuration de mesures réglementaires correspondant aux dates souhaitées

Etape 11-15 : Le job Talend *generation_E2* collecte toutes les données horaires ainsi que les données référentielles associées depuis les bases de données MongoDB et PostgreSQL et génère le dataset E2 complet au format XML à partir des données collectées

Etape 16-18 : S'il s'agit du mode « auto », le job Talend *generation_E2* filtre parmi les données récupérées aux étapes 11-15, les données horaires des SPO ayant été mises à jour depuis la dernière génération (nouvelle mesure ou modification d'une valeur). Le dataset E2 partiel au format XML est ensuite généré selon les filtres sur le réseaux de mesure et le type de la donnée (brute ou validée)

Etape 19-20 : Le job Talend *generation_E2* met à jour les données de suivi de génération des datasets E2 (base postgresQL)

Etape 21 : Le job Talend *generation_E2* dépose les datasets générés dans les répertoires cibles

Etape 22 : Fin du job Talend

Spécifications fonctionnelles détaillées

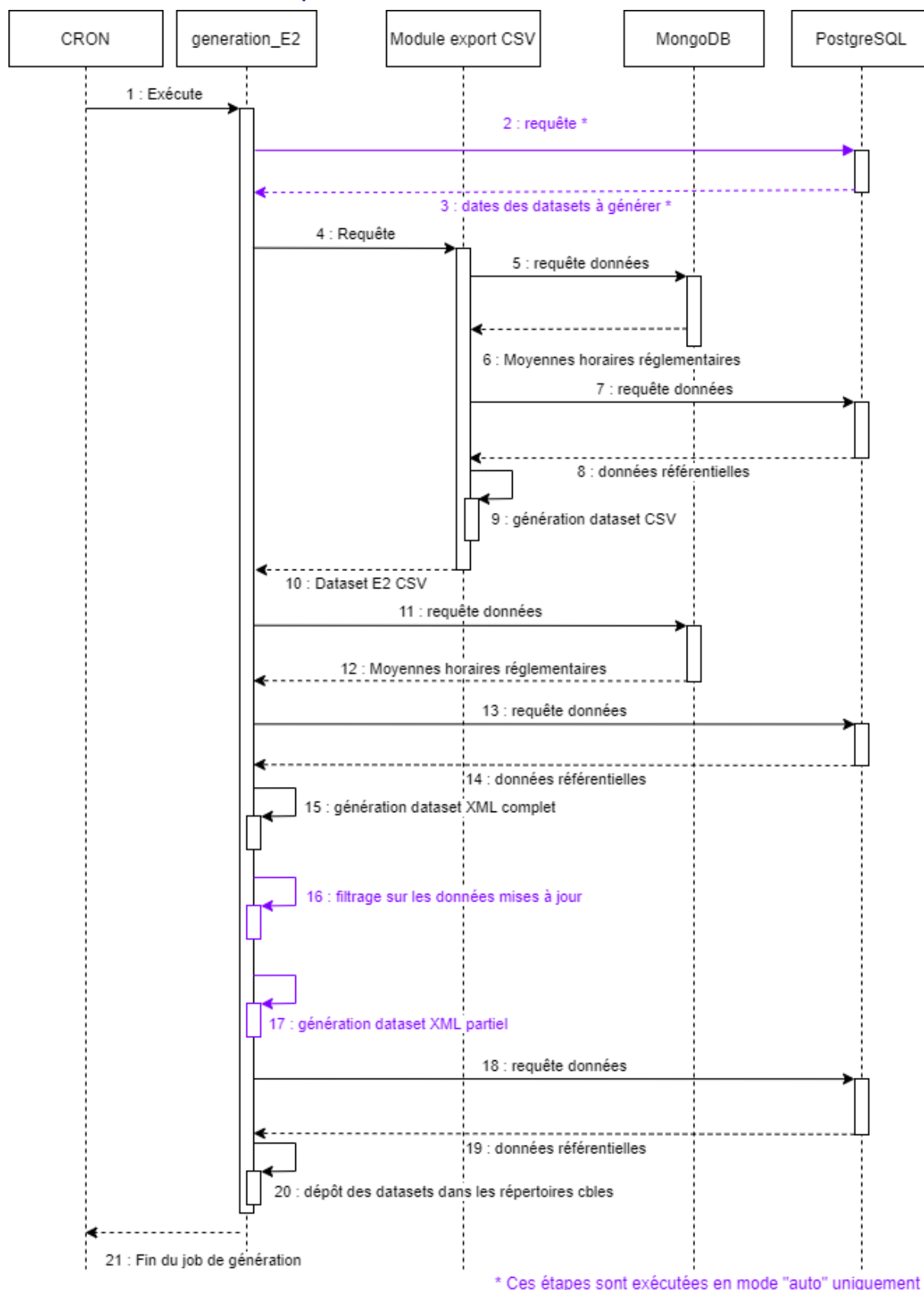


Figure 4 : diagramme de séquence de la génération des datasets E2

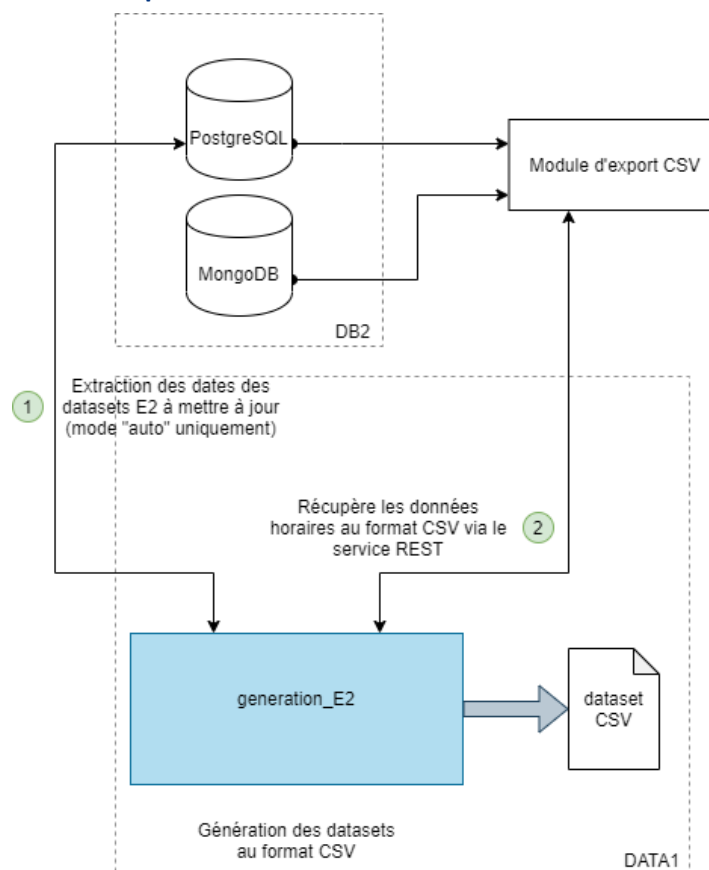


Figure 5 : Interaction entre les modules impliqués dans la génération des datasets E2 au format CSV

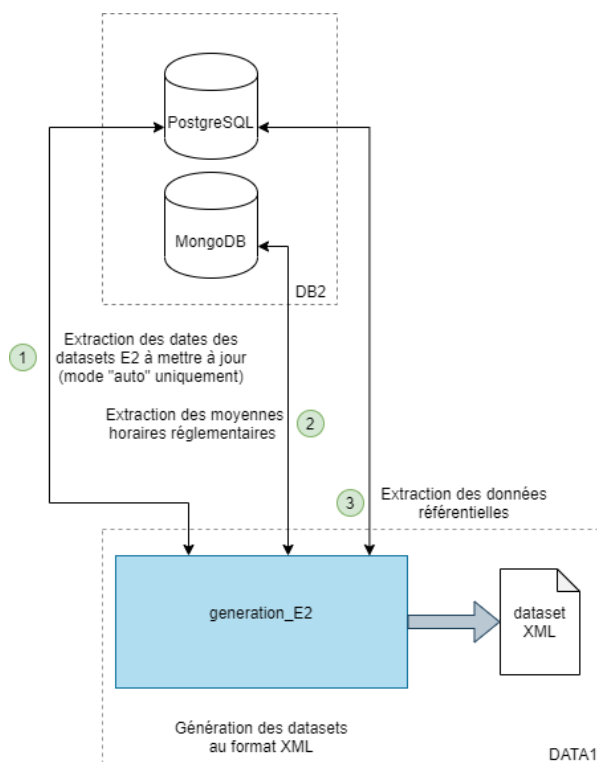


Figure 6 : Interaction entre les modules impliqués dans la génération des datasets E2 au format XML

Spécifications fonctionnelles détaillées

Désignation	Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
Exigences	EX-0020 EX-0040 EX-0050 EX-0060 EX-0070 EX-0080 EX-0090 EX-00100 EX-0120 EX-0130 EX-0140
Accès	Job Talend exécutable depuis DATA1
Description	<p>Ce job est chargé de générer les datasets E2 complet aux formats CSV, XML complet et XML partiel pour un ensemble de dates données.</p> <p>Le job peut-être exécuté selon deux modes : « auto » ou « manu ».</p> <p>En mode « manu » seuls les datasets E2 complet sont générés.</p> <p>Les diagrammes Figure 5 et Figure 6 illustrent les interactions du job Talend avec les autres modules de GEOD'AIR dans le processus de génération des datasets E2.</p>
Entrées	<p>Les dates nécessitant une génération de dataset E2 sont collectées différemment selon le mode d'exécution :</p> <ul style="list-style-type: none"> Mode « auto » : Le job extrait la liste des datasets à générer depuis une table PostgreSQL dédiée au suivi de la génération des datasets E2 <i>rapport_ue</i> et <i>rapport_historique</i>. Cf. [SP-F-0010] – Collecter les dates des datasets E2 à mettre à jour au fil de l'eau Mode « manu » : Les dates sont fournies en paramètre du job
Paramètres	<p>Arguments obligatoires :</p> <ul style="list-style-type: none"> Mode de génération manuel : booléen <i>true</i> pour le mode « manu », <i>false</i> sinon Chemin du fichier de configuration : cf. [SP-F-0050] – Paramétrer le job Talend generation_E2 <p>Arguments obligatoires en mode manuel uniquement :</p> <ul style="list-style-type: none"> Date de début de la période au format aaaa-MM-jj Date de fin de la période au format aaaa-MM-jj <p>Arguments optionnels :</p> <ul style="list-style-type: none"> Transfert vers data.gouv.fr : booléen indiquant si les datasets E2 doivent être transférés vers data.gouv.fr Transfert vers l'agence européenne : booléen indiquant si les datasets E2 doivent être transférés vers l'agence européenne
Traitements	<p>Cf. Figure 4 :</p> <ul style="list-style-type: none"> Détection du mode (« auto » ou « manu ») et contrôle des données d'entrée Chargement du contexte du job à partir du fichier de configuration Collecte des dates pour lesquelles une régénération des datasets E2 est requise : <ul style="list-style-type: none"> En mode « manu » : ensemble des dates incluses dans la période renseignées en entrée du job

Spécifications fonctionnelles détaillées

	<ul style="list-style-type: none"> ○ En mode « auto » : dans les tables postgresSQL dédiée, toutes les date pour lesquelles le dataset E2 à le champ « statut » à la valeur « En attente » sont collectées <p>Ces étapes sont exécutées pour chacune des dates collectées :</p> <ul style="list-style-type: none"> • Téléchargement du dataset E2 complet au format CSV via un appel au service REST du module d'export CSV : cf. [SP-F-0060] – Générer un dataset E2 . Cette requête est faite au travers dun composant Talend tHttpRequest • Extraction de l'ensemble des moyennes horaires réglementaires sur la journée demandée depuis la base de données StatsPeriodiques de MongoDB ainsi que des données référentielles nécessaires à la génération du dataset E2 complet au format XML • Génération du dataset E2 complet au format XML à partir des données extraites • Filtrage des moyennes horaires réglementaires des SPO ayant été mises à jour depuis la dernière génération sur la journée demandée, depuis les données récupérées pour les datasets E2 complets au format XML • Filtrage sur les réseaux de mesures présent dans l'ensemble des moyennes horaires réglementaires des SPO précédemment récupérées • Étapes exécutées pour chaque réseaux de mesures identifiés : <ul style="list-style-type: none"> ○ Filtrage sur le type de la données (brute ou validée) des moyennes horaires réglementaires des SPO ○ Génération du dataset E2 partiel au format XML à partir des données filtrées • Mise à jour des tables PostgreSQL de suivi des datasets E2, si la génération des datasets CSV et XML s'est déroulée sans erreur : cf. §Sortie ci-dessous • Dépôt des fichiers générés (XML et CSV) dans les répertoires cibles d'envoi. Le fichier déposé écrase le dernier de même nom si ce dernier existe déjà : <ul style="list-style-type: none"> ○ Si l'option de transfert vers data.gouv.fr est demandée, seuls les datasets <u>complet (XML et CSV)</u> sont déposés dans le répertoire d'envoi vers data.gouv.fr ○ Si l'option de transfert vers l'agence européenne est demandée, les datasets <u>partiels au format XML uniquement</u> sont déposés dans le répertoire d'envoi vers l'Agence Européenne ○ Si aucune des deux options n'est demandée, tous les datasets sont déposés dans un répertoire dédié au stockage. Ces datasets ne seront pas transférés.
Sorties	<p>Les fichiers générés respectent la convention de nommage décrite au §3.1.4. Le format XML du dataset E2 est défini dans le document [REF-0030].</p> <p>La table <i>rapport_ue</i> est mise à jour avec les champs suivants :</p> <ul style="list-style-type: none"> • Statut : « Terminé » • Date_creation : date courante <p>La table <i>rapport_historique</i> est mis à jour avec les champs suivants :</p> <ul style="list-style-type: none"> • Date_création : date courante

4.1.5. [SP-F-0050] – Paramétrer le job Talend generation_E2

Désignation	Paramétrer le job generation_E2
Exigences	EX-0050
Accès	Fichier de configuration sur DATA1
Description	Le job de génération des datasets E2 <i>generation_E2</i> est paramétrable au travers du fichier de configuration <i>e2Conf.ini</i> . NB : Ce fichier de configuration est commun aux différents scripts impliqués dans la génération, la mise à jour et le transfert des datasets E2.
Entrées	NA
Paramètres	Paramètres renseignés dans le fichier de configuration : <ul style="list-style-type: none"> • Chemin du répertoire d'envoi des datasets E2 vers l'agence européenne • Chemin du répertoire d'envoi des datasets E2 vers data.gouv.fr • Chemin du répertoire de stockage par défaut des datasets E2 • Chemin du fichier de logs du job <i>generation_E2</i>
Traitements	Les paramètres consignés dans ce fichier de configuration s'appliquent au job Talend <i>generation_E2</i> : cf. [SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2 Les différents paramètres sont traduits en variables de contexte Talend. Le job Talend possède une étape de lecture du fichier de configuration et de chargement dynamique du contexte (voir composant Talend tContextLoader).
Sorties	NA

4.1.6. [SP-F-0060] – Générer un dataset E2 via l'API REST du module d'export CSV

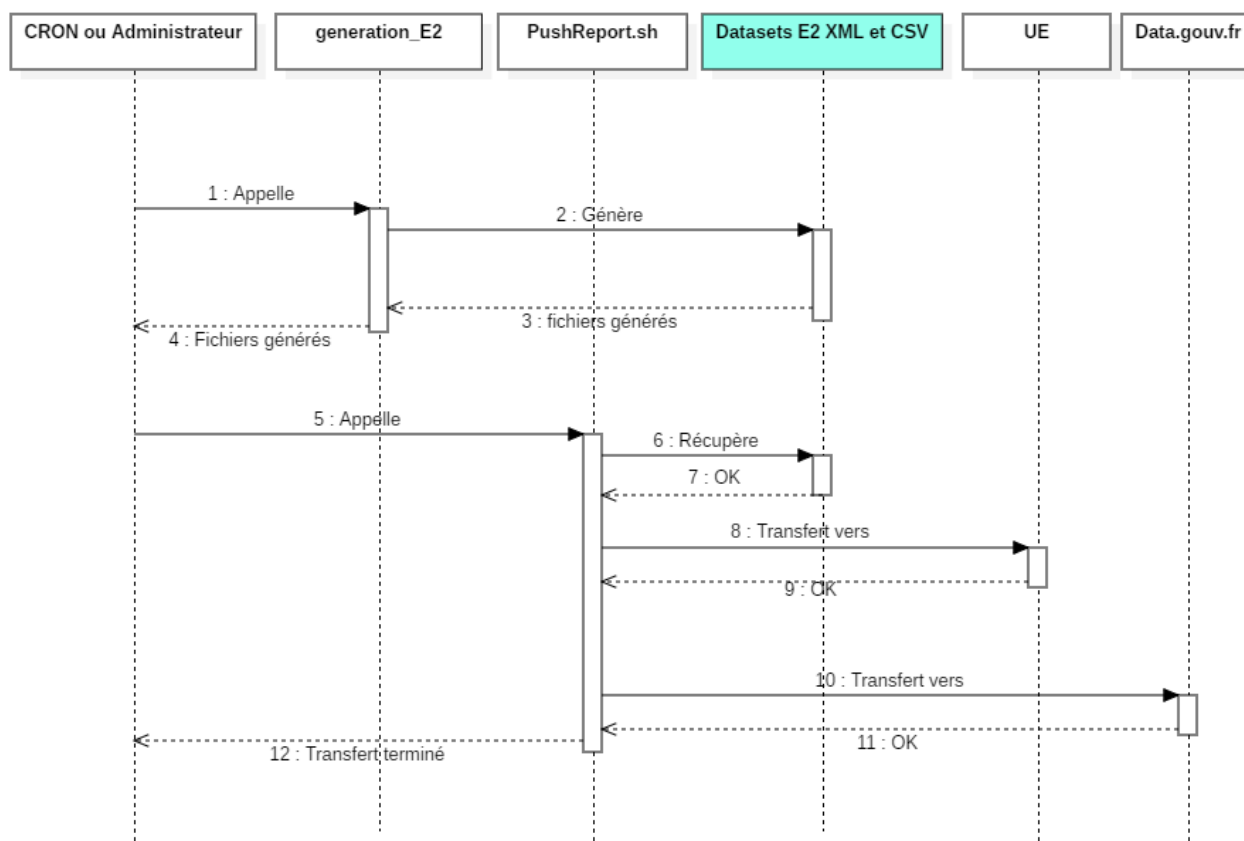
Désignation	Générer un dataset E2 au format CSV via un service REST PHP
Exigences	EX-0110
Accès	API REST accessible depuis DATA1
Description	Ce service est chargé d'exporter les moyennes horaires réglementaires intégrées dans MongoDB sous forme d'un dataset E2 au format CSV. Le format des données téléchargées via ce service est identique au format d'export via l'IHM.
Entrées	<ul style="list-style-type: none"> • Date du dataset E2 au format « yyyy-MM-jj »
Paramètres	NA
Traitements	<ul style="list-style-type: none"> • Extraction de l'ensemble des moyennes horaires issues de configuration de mesures réglementaires sur la journée demandée depuis la base de données StatsPeriodiques de MongoDB • Génération du fichier CSV dans le même format que celui utilisé pour exporter les données depuis l'IHM d'accès aux données
Sorties	Téléchargement du fichier CSV généré. Le fichier généré est le dataset E2 au format CSV, il respecte la convention de nommage décrite au §3.1.4

4.2. Mise à disposition des datasets E2

Le diagramme de séquence ci-dessous illustre le fonctionnement de la génération des datasets E2 et de leur transfert vers les serveurs cibles data.gouv.fr et l'Agence Européenne.

Les acteurs et modules intervenants dans ces processus sont les suivants :

- CRON : gestionnaire de tâches périodique (cf. [SP-F-0030] – Générer périodiquement des datasets E2 de manière automatique)
- Administrateur : utilisateur GEOD'AIR ayant les droits de connexions sur DATA1 et pouvant par conséquent faire appel au script de génération des datasets E2 (cf. [SP-F-0020] – Générer manuellement des datasets E2 via appel au script GenerationE2.sh)
- « generation_E2 » : cf. [SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
- « PushReport.sh » : script bash chargé de l'acheminement des datasets E2 vers les serveurs cibles
- « UE » : serveur FTP de l'Agence Européenne
- « Data.gouv.fr » : serveur FTP de data.gouv



Les étapes du processus sont les suivantes :

Etape 1 : Le job Talend generation_E2 est appelé soit manuellement par un administrateur, soit de manière automatique par une tâche CRON

Etape 2 : Le job Talend generation_E2 crée les fichiers au format CSV et XML. Ces fichiers sont déposés dans un répertoire dédié sur DATA1

Etape 3-4 : La génération est terminée : Fin du job de génération des datasets E2

Etape 5 : Le script PushReport.sh est appelé soit manuellement par un administrateur, soit de manière automatique par une tâche CRON

Etape 6-7 : Le script PushReport.sh récupère les datasets en attente de transfert

Etape 8-9 : Le script PushReport.sh transfère les datasets vers l'Agence Européenne

Etape 10-11 : Le script PushReport.sh transfère les datasets vers data.gouv.fr

Etape 12 : Le transfert est terminé : fin du script PushReport.sh

4.2.1. [SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh

Désignation	Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh
Exigences	EX-0160 EX-0190 EX-0210
Accès	Commande bash exécutable depuis DATA1
Description	Le script bash PushReport.sh est exécutable manuellement via une commande depuis un terminal. Il est chargé d'envoyer les datasets E2 générés vers les serveurs de data.gouv.fr et de l'Agence Européenne
Entrées	Ensemble des fichiers datasets E2 présents dans les répertoires d'envoi vers l'Agence Européenne et vers data.gouv.fr
Paramètres	<p>Arguments optionnels :</p> <p><u>Choix du (ou des) serveur(s) cible(s) :</u></p> <ul style="list-style-type: none"> --datagouv : seuls les datasets destinés à l'envoi vers data.gouv.fr seront transférés --ue : seuls les datasets destinés à l'envoi vers l'Agence Européenne seront transférés <p><u>Choix des fichiers à envoyer (en fonction de la date des données) :</u></p> <ul style="list-style-type: none"> --current-day : seuls les fichiers dont les données sont datées du jour J sont transférés --previous-day : seuls les fichiers dont les données sont datées du jour J-1 sont transférés --last2weeks : seuls les fichiers dont les données sont datées des deux dernières semaines sont transférés --older : seuls les fichiers dont les données sont datées d'au-delà des deux dernières semaines sont transférés
Traitements	<ul style="list-style-type: none"> Transfert des datasets E2 présents dans le répertoire d'envoi sur le serveur de data.gouv.fr : cf. [SP-F-0100] – Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr (ce transfert est exécuté par le script pushDataGouv.sh) Transfert des datasets E2 présents dans le répertoire d'envoi sur le serveur de l'Agence Européenne : cf. [SP-F-0110] – Transférer les datasets E2 sur le serveur FTP de l'Agence Européenne Archivage des datasets E2 envoyés
Sorties	Les logs d'exécution du scripts sont consignés dans un fichier de sortie dont le chemin est défini dans le fichier de configuration du script.

4.2.2. [SP-F-0080] – Transférer périodiquement les datasets E2 vers les serveurs cibles

Désignation	Générer périodiquement des datasets E2 de manière automatique
Exigences	EX-0160 EX-0190
Accès	Tâche exécutée automatiquement de manière périodique
Description	<p>Pour le transfert vers l'Agence européenne, le script de transfert des datasets E2 (cf.[SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh) est appelé dans le script Generation_E2_auto.sh qui est exécuté au travers d'une tâche CRON.</p> <p>Les fichiers E2 sont envoyés toutes les heures à XXh15min.</p> <p><u>Exécution du script avec les paramètres de PushReport.sh suivants :</u></p> <ul style="list-style-type: none"> - envoi des datasets E2 du jour J (paramètres <code>--ue --current-day</code>) - envoi des datasets E2 du jour J-1 (paramètres <code>--ue --previous-day</code>) - envoi des datasets E2 des deux dernières semaines (paramètres <code>--ue --last2weeks</code>) - envoi des datasets E2 au-delà des deux dernières semaines (paramètres <code>--ue --older</code>) <p>Pour le transfert vers datagouv, le script de transfert des datasets E2 (cf.[SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh) est exécuté périodiquement au travers d'une tâche CRON.</p> <p>Les fichiers E2 sont envoyés toutes les heures à XXh30min : exécution du script avec le paramètre <code>--datagouv</code></p>
Entrées	Paramétrage du script : cf.[SP-F-0090] – Configurer le script PushReport.sh
Paramètres	NA
Traitements	Exécution du script PushReport.sh avec les paramètres fournis dans la description
Sorties	NA

4.2.3. [SP-F-0090] – Configurer le script PushReport.sh

Désignation	Configurer les répertoires de dépôt sur le serveur FTP de Data.gouv.fr
Exigences	EX-0220
Accès	Fichier de configuration sur DATA1
Description	<p>Le script bash PushReport.sh est paramétrable au travers du fichier de configuration <code>e2Conf.ini</code>. Il permet notamment de configurer les chemins des répertoire de dépôt des datasets sur le serveur FTP de data.gouv.fr</p> <p>NB : Ce fichier de configuration est commun aux différents scripts impliqués dans la génération, la mise à jour et le transfert des datasets E2.</p> <p>Par ailleurs, un second fichier de configuration externe à l'image contient les paramètres éditables par l'INERIS.</p>
Entrées	NA

Spécifications fonctionnelles détaillées

Paramètres	<p>Paramètres renseignés dans le fichier de configuration <i>e2Conf.ini</i> :</p> <ul style="list-style-type: none"> • Chemin du répertoire d'envoi des datasets E2 vers l'agence européenne • Chemin du répertoire d'envoi des datasets E2 vers data.gouv.fr • Chemin du fichier de logs du script <i>PushReport.sh</i> <p>Paramètres éditables par l'INERIS :</p> <ul style="list-style-type: none"> • Chemin du répertoire d'archive des datasets envoyés à l'Agence européenne • Chemin du répertoire d'archive des datasets envoyés à data.gouv.fr (NB : le répertoire d'archive peut être commun aux datasets envoyés à l'Agence Européenne) • Chemin d'accès au répertoire de dépôt des données sur le FTP de data.gouv.fr
Traitements	Les paramètres consignés dans ce fichiers de configuration s'appliquent au script <i>PushReport.sh</i> : cf. [SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles
Sorties	NA

4.2.4. [SP-F-0100] – Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr via le script *pushDataGouv.sh*

Désignation	Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr via le script <i>pushDataGouv.sh</i>
Exigences	EX-0200 EX-0230 EX-0240 EX-0250
Accès	Commande bash exécutable depuis DATA1
Description	Le script bash de transfert des datasets E2 vers le serveur FTP de data.gouv.fr est exécutable manuellement via une commande depuis un terminal. Il est également appelé par le script <i>PushReport.sh</i> .
Entrées	Ensemble des fichiers datasets E2 présents dans le répertoire d'envoi vers data.gouv.fr
Paramètres	Paramétrage du script : cf. [SP-F-0090] – Configurer le script <i>PushReport.sh</i>
Traitements	<p>Fichiers transférés :</p> <p>Les fichiers E2 envoyés à data.gouv pour une date donnée sont les suivants :</p> <ul style="list-style-type: none"> • Dataset E2 complet au format XML • Dataset E2 complet au format CSV <p>Répertoires de destination :</p> <p>Le répertoire de dépôt est spécifié dans le fichier de configuration</p> <p>Etapes du transfert :</p> <ul style="list-style-type: none"> • Connexion au serveur FTP • Création d'un répertoire portant comme nom l'année du dataset au format « aaaa » si ce dernier n'existe pas encore dans le répertoire de destination • Dépôt des fichiers sur le serveur, dans le répertoire correspondant à l'année du dataset. Le dataset E2 écrase le dernier fichier de même nom si celui-ci existe déjà • Archivage du dataset sur DATA1 dans le répertoire d'archive spécifié dans le fichier de configuration

Sorties	Les logs d'exécution du scripts sont consignés dans un fichier de sortie dont le chemin est défini dans le fichier de configuration du script PushReport.sh.
----------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------

4.2.5. [SP-F-0110] – Transférer les datasets E2 sur le serveur FTP de l'Agence Européenne

Désignation	Transférer les datasets E2 sur le serveur FTP de l'Agence Européenne
Exigences	EX-0140 EX-0170
Accès	Opération exécutée par le script PushReport.sh : cf. [SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh
Description	Opération de transfert des datasets E2 vers le serveurs FTP de l'Agence Européenne
Entrées	Ensemble des fichiers datasets E2 partiels présents dans le répertoire d'envoi vers l'Agence Européenne
Paramètres	Paramétrage : cf. [SP-F-0090] – Configurer le script PushReport.sh
Traitements	<p>Fichiers transférés : Les fichiers E2 envoyés à l'Agence Européenne pour une date donnée sont les datasets E2 partiels au format XML uniquement. NB : les datasets E2 de format CSV ne sont pas transférés à l'Agence Européenne</p> <p>Etapes du transfert :</p> <ul style="list-style-type: none"> • Connexion au serveur FTP • Dépôt des fichiers sur le serveur. Le dataset E2 écrase le dernier fichier de même nom si celui-ci existe déjà • Archivage du dataset sur DATA1 dans le répertoire d'archive spécifié dans le fichier de configuration
Sorties	Les logs d'exécution du scripts sont consignés dans un fichier de sortie dont le chemin est défini dans le fichier de configuration du script PushReport.sh.

4.2.6. [SP-F-0120] – Alerter l'administrateur en cas d'échec de transfert

Désignation	Alerter l'administrateur en cas d'échec de transfert
Exigences	EX-0180 EX-0260
Accès	Opération exécutée par Zabbix
Description	Alerte par email en cas d'échec de transfert d'un ou plusieurs datasets E2 vers l'un des serveurs cibles (data.gouv.fr ou Agence Européenne)
Entrées	Ensemble des fichiers datasets E2 présents dans le répertoire d'envoi vers l'Agence Européenne ou data.gouv.fr
Paramètres	NA
Traitements	Un agent zabbix est chargé de détecter les éventuelles erreurs de transferts vers les serveurs cibles. En cas d'échec d'un transfert, les administrateur de GEOD'AIR sont notifiés par email. L'email contient la liste des fichiers non transférés.
Sorties	NA

4.2.7. [SP-F-0130] – Mettre à jour l'e-mail de rapport quotidien

Désignation	Mettre à jour l'e-mail de rapport quotidien
--------------------	---------------------------------------------

CONFIDENTIEL

Spécifications fonctionnelles détaillées

Exigences	EX-0270
Accès	Tâche exécutée automatiquement de manière quotidienne. Opération exécutée par le script parseLog.sh de manière quotidienne via un cron.
Description	Le script d'envoi du rapport d'intégration ETL, parseLog.sh est exécuté quotidiennement au travers d'une tâche CRON.
Entrées	Ensemble des statistiques d'intégration ETL sur la journée J-1. Des statistiques liées à la consultation du site GP
Paramètres	NA
Traitements	Exécution du script parseLog.sh
Sorties	<p>Contenu de l'e-mail :</p> <p>Statistiques JOUR: JJ/MM/AA Fichiers integres: XXXX Fichiers integres OK: XXXX Fichiers integres KO: XXXX Fichiers T integres: XXXX Fichiers V integres: XXXX Rapports E2 XML transférés vers data.gouv.fr : XXXX Rapports E2 CSV transférés vers data.gouv.fr: XXXX Rapports E2 XML transférés vers l'UE: XXXX Export avancé API MoyH : XXXX Export avancé API MoyJ : XXXX Export avancé API MoyJH : XXXX Export avancé API MoyA : XXXX Export avancé API Pointdeprelevement : XXXX Export avancé API station : XXXX Nombre d'utilisateurs avancés : XXXX Export basique API MoyH : XXXX Export basique API MoyJ : XXXX Export basique API MoyJH : XXXX Export basique API MoyA : XXXX Export basique API Pointdeprelevement : XXXX Export basique API station : XXXX Nombre d'utilisateurs basique : XXXX</p>

5. Couverture des exigences

Réf.	Exigence	Spécifications
EX-0010	Les fichiers E2 nécessitant une mise à jour sont générés toutes les heures de manière automatique.	[SP-F-0010] – Collecter les dates des datasets E2 à mettre à jour au fil de l'eau [SP-F-0030] – Générer périodiquement des datasets E2 de manière automatique
EX-0020	En mode automatique, les fichiers E2 générés pour une date donnée sont les suivants : <ul style="list-style-type: none"> Dataset E2 complet au format XML Dataset E2 complet au format CSV Dataset E2 partiel au format XML : contient uniquement les moyennes horaires des SPO ayant été mises à jour depuis la dernière génération (nouvelle mesure ou modification d'une valeur). Les datasets E2 partiels ne sont transmis qu'à l'agence Européenne. 	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
EX-0030	Les fichiers E2 peuvent être générés manuellement à partir d'une période temporelle antérieure via une commande. La période est renseignée par l'utilisateur par une date de début et date de fin (dates incluses dans la période).	[SP-F-0020] – Générer manuellement des datasets E2 via appel au script GenerationE2.sh
EX-0040	En mode manuel, les fichiers E2 générés pour une date donnée sont les suivants : <ul style="list-style-type: none"> Dataset E2 complet au format XML Dataset E2 complet au format CSV 	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
EX-0050	Le script de génération manuelle des datasets E2 possède les options suivantes : <ul style="list-style-type: none"> Une option de diffusion vers data.gouv.fr Une option de diffusion vers l'Agence Européenne Si aucune des deux options n'est renseignée par l'utilisateur dans la commande, les datasets générés ne seront pas envoyés. Ils sont stockés dans un répertoire dédié.	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2 [SP-F-0050] – Paramétrer le job Talend generation_E2
EX-0060	Un dataset E2 <u>complet</u> rapporte les moyennes horaires réglementaires (brutes et validées) de la journée de l'ensemble des points de prélèvements (SPO) des régions surveillées par le dispositif.	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job
EX-0070	Un dataset E2 <u>partiel</u> rapporte uniquement les moyennes horaires réglementaires (brutes ou validées) de la journée et dont la date d'insertion (ou de mise à jour) dans Geod'Air est postérieure à la date de génération du dernier dataset E2. Si aucun dataset E2 partiel n'a encore été généré pour la journée, il faut prendre en compte les moyennes établies depuis 00h00.	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job
EX-0080	L'ensemble des données contenu dans un dataset E2 <u>partiel</u> ne concerne qu'un unique réseau de mesure.	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2

EX-0090	L'ensemble des données contenu dans un dataset E2 <u>partiel</u> ne concerne qu'un seul type de données (brutes ou validées).	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
EX-0100	Dans un dataset E2 XML, la distinction entre les données brutes et validées est portée uniquement par le flag observationverification (http://dd.eionet.europa.eu/vocabulary/aq/observationverification) du bloc <swe:values>. Celui-ci permet de classifier dans le fichier E2 chaque moyenne horaire transmise : « 3 » pour les données brutes et « 2 » pour les données vérifiées (validées).	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
EX-0110	Le format de données du dataset CSV est exactement identique à celui d'un export CSV depuis l'IHM d'accès aux données de GEOD'AIR.	[SP-F-0060] – Générer un dataset E2 via l'API REST du module d'export CSV
EX-0120	Un dataset E2 <u>complet</u> est nommé de la manière suivante : FR_E2_AAAA-MM-JJ.YYY : <ul style="list-style-type: none"> • AAAA : représente l'année des données présentes dans le fichier, • MM : représente le numéro du mois des données présentes dans le fichier, • JJ : représente le numéro du jour dans le mois des données présentes dans le fichier, • YYY : extension relative au format de fichier « xml » ou « csv » 	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
EX-0130	Un dataset E2 <u>partiel</u> est nommé de la manière suivante : FR_E2_P_AAAA-MM-JJ-hhmmss.YYY : <ul style="list-style-type: none"> • AAAA : représente l'année des données présentes dans le fichier, • MM : représente le numéro du mois des données présentes dans le fichier, • JJ : représente le numéro du jour dans le mois des données présentes dans le fichier, • hhmmss : représente l'heure de génération du rapport • YYY : extension relative au format de fichier « xml » ou « csv » <p>La lettre « P » est rajoutée entre la mention « E2 » et l'année pour indiquer qu'il s'agit d'un dataset partiel.</p>	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
EX-0140	Lorsqu'un dataset E2 est généré par GEOD'AIR, il écrase le dernier dataset de même nom généré précédemment, si ce dernier existe déjà.	[SP-F-0040] – Générer les datasets E2 au format CSV et XML à l'aide du job generation_E2
EX-0150	Un dataset E2 est mis à jour uniquement dans l'une des conditions suivantes : <ul style="list-style-type: none"> • Les moyennes horaires rapportées dans le dernier dataset E2 ont été mises à jour via le flux d'intégration des données automatiques 	[SP-F-0010] – Collecter les dates des datasets E2 à mettre à jour au fil de l'eau

	<ul style="list-style-type: none"> De nouvelles moyennes horaires ont été intégrées et n'ont pas encore été rapportées dans le dernier dataset E2 généré <p>NB : En cas de reprise des données, le dataset E2 n'est pas régénéré.</p>	
EX-0160	Les datasets E2 ayant été générés manuellement avec l'option de diffusion vers l'Agence européenne ou bien automatiquement par Geod'air sont déposés sur le serveur FTP dédié de l'Agence Européenne. Toutes les heures le transfert des E2 s'effectue quel que soit l'ancienneté des données.	<p>[SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh</p> <p>[SP-F-0080] – Transférer périodiquement les datasets E2 vers les serveurs cibles</p>
EX-0170	Tous les fichiers E2 <u>partiels</u> au format XML générés depuis le dernier envoi sont envoyés à l'Agence Européenne. N.B. Les fichiers E2 <u>complets</u> ne sont pas envoyés.	[SP-F-0110] – Transférer les datasets E2 sur le serveur FTP de l'Agence Européenne
EX-0180	Un mail est envoyé à l'administrateur GEOD'AIR dans le cas où l'application GEOD'AIR n'est en possibilité de pousser les fichiers E2 vers l'Agence Européenne.	[SP-F-0120] – Alerter l'administrateur en cas d'échec de transfert
EX-0190	Toutes les heures, tous les datasets E2 <u>complets</u> , qu'ils aient été générés manuellement ou bien automatiquement par GEOD'AIR, sont déposés sur le serveur FTP dédié de data.gouv.fr.	<p>[SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh</p> <p>[SP-F-0080] – Transférer périodiquement les datasets E2 vers les serveurs cibles</p>
EX-0200	Seuls les fichiers E2 <u>complets</u> au format XML et CSV sont envoyés à data.gouv.	[SP-F-0100] – Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr via le script pushDataGouv.sh
EX-0210	Seule la version la plus récente du dataset E2 est envoyé à data.gouv.fr.	[SP-F-0070] – Transférer les datasets E2 vers les serveurs cibles via appel au script PushReport.sh
EX-0220	Le chemin d'accès du répertoire de dépôt des datasets E2 sur le serveur FTP est configurable depuis un fichier de configuration qui peut être modifié à tout moment par l'INERIS.	[SP-F-0090] – Configurer le script PushReport.sh
EX-0230	Au sein du répertoire de dépôt des datasets E2 cité précédemment, les fichiers sont triés dans des répertoires par année. Le format de nommage de ces sous-répertoires est « AAAA ».	[SP-F-0100] – Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr via le script pushDataGouv.sh
EX-0240	Lorsqu'un fichier est déposé sur le serveur FTP de data.gouv.fr, celui-ci est classé dans le sous-répertoire correspondant à son année. Si le répertoire de l'année n'existe pas, il est automatiquement créé et le fichier y est déposé.	[SP-F-0100] – Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr via le script pushDataGouv.sh

Spécifications fonctionnelles détaillées

EX-0250	Lorsqu'un dataset E2 est déposé sur data.gouv.fr, il écrase le dernier dataset E2 du même nom.	[SP-F-0100] – Transférer les dataset E2 sur le serveur FTP de Data.gouv.fr via le script pushDataGouv.sh
EX-0260	Un mail est envoyé à l'administrateur GEOD'AIR dans le cas où l'application GEOD'AIR n'est pas en possibilité de pousser les fichiers E2 vers data.gouv.	[SP-F-0120] – Alerter l'administrateur en cas d'échec de transfert
EX-0270	Le mail de rapport intégration ETL ne change pas. La ligne « Rapports E2 XML transférés vers l'UE » concerne bien le nombre de datasets E2 <u>partiels</u> .	[SP-F-0130] – Mettre à jour l'e-mail de rapport quotidien

6. Les documents en référence

6.1. Fourniture INERIS

[REF-0010]

Conception Détaillée - flux d'acquisition et production de données, BULL

[REF-0020]

Mise à jour du flux E2 vers agence européenne et data.gouv.fr - Spécifications fonctionnelles

ref. : 2020-04 - Ajout évolution data.gouv_v1.1.docx

[REF-0030]

USER GUIDE TO XML & DATA MODEL-v3.4, Jaume Targa and Tony Bush (ETC/ACM)

UserGuide2_AQD_XML_v3.4.1.pdf

6.2. Fourniture IT Link

[REF-0040]

Devis - Mise à jour du flux E2 vers agence européenne et data.gouv.fr, IT LINK,

ref. : GEODAIR-TMAE014_Evolution export data.gouv-Version allégée-SFD01-1.00.pdf

[REF-0050]

Mise à jour du flux E2 vers agence européenne et data.gouv.fr - Reformulation du besoin & définition des exigences

ref. : Export des datasets E2-Exigences-4.00.pdf