
NOTE POUR LE GROUPE DE PILOTAGE SANDRE

Réflexions sur la gestion du différentiel d'un jeu de données

Titre : Réflexions sur la gestion du différentiel d'un jeu de données

Créateur : Système d'Information sur l'Eau / Sandre

Éditeur : Ministère en charge de l'écologie et du développement durable

Contributeur : Office International de l'Eau / Sandre

Auteur : Dimitri Meunier

Résumé : Cette note présente la manière dont sont gérés les différents changements d'état d'un jeu de données au sein d'un scénario d'échange Sandre.

Date : 07/03/2016

Version : 1

Type : Texte

Format : PDF

Identifiant : urn:sandre:note-methodologique:DifferentielJDD::::ressource:1:::pdf

Langue : fra

Couverture spatiale : France métropolitaine

Couverture temporelle :

Droits d'usage : <http://creativecommons.org/licenses/by-nc-sa/2.0/fr>

Table des matières

1.. Contexte.....	2
2.. Scénarios d'échanges concernés	3
2.1Qualité des eaux superficielles continentales version 3	4
2.2Quantité et qualité des eaux souterraines version 2	5
3.. Attributs date de création, de mise à jour des objets.....	7
4.. Scénario de service web concerné	7
5.. Perspectives.....	8

📌 1. Contexte

Au titre du [Schéma National des Données sur l'Eau \(SNDE\)](#) (cf. [l'arrêté interministériel publié au JO du 24 août 2010](#)), pour satisfaire à des besoins de connaissance, réglementaires, de qualité, etc, les organismes qui produisent (i.e. producteurs de données) des jeux de données pour le Système d'Information sur l'Eau (SIE) doivent les échanger selon les spécifications Sandre. Le Service d'administration nationale des données et des référentiels sur l'eau (Sandre) diffuse librement les spécifications d'échanges de données et les jeux de données de référence pour SIE. L'ensemble des jeux de données, des jeux de données de référence et des spécifications est diffusé par les sites web accessibles par le [portail Web Eaufrance](#).

Par nature, un jeu de données peut évoluer dans le temps. Prenons l'exemple d'une fleur (i.e. objet), en la regardant pendant un moment (voir dessin ci-dessous), on observe des changements (sa taille, son nombre de feuilles, son nombre de pétales, sa maturité...). La situation dans laquelle se trouve un objet à une date donnée représente son état. A chaque état correspond un jeu de données. Un producteur de données outillé est capable de restituer, à la demande de l'utilisateur, le dernier état d'un objet voire les états d'un objet et les différences entre eux. L'utilisateur peut ainsi mieux analyser ces données d'observation.



-----> Temps

Souhaitant informer l'utilisateur qu'un objet a changé d'état depuis sa dernière diffusion, le Sandre a ajouté - à titre d'expérimentation et à la demande des producteurs de données concernés - l'attribut *Action* dans les dernières versions des scénarios Sandre pour l'échange et la diffusion au format XML de jeux de données sur :

- la qualité et quantité des eaux souterraines,
- la qualité des eaux superficielles continentales.

Parallèlement, voulant informer l'utilisateur des raisons pour lesquelles un objet a changé d'état depuis sa dernière diffusion, le Sandre a conceptualisé la notion de généalogie. Elle permet de tracer les modifications apportées à chaque élément d'un référentiel administré et/ou diffusé par le Sandre. Le changement d'état d'un objet de référence est donc conservé selon 3 informations : la date de changement d'état, la raison de l'évolution (cf. [nomenclature n°695](#)) et d'éventuelles précisions. Cette notion répond à une notion UML définie par le Sandre (cf. image n°0). Néanmoins, elle est limitée aux jeux de données de référence et elle ne permet pas de gérer les évolutions au niveau de chaque attribut de l'objet.

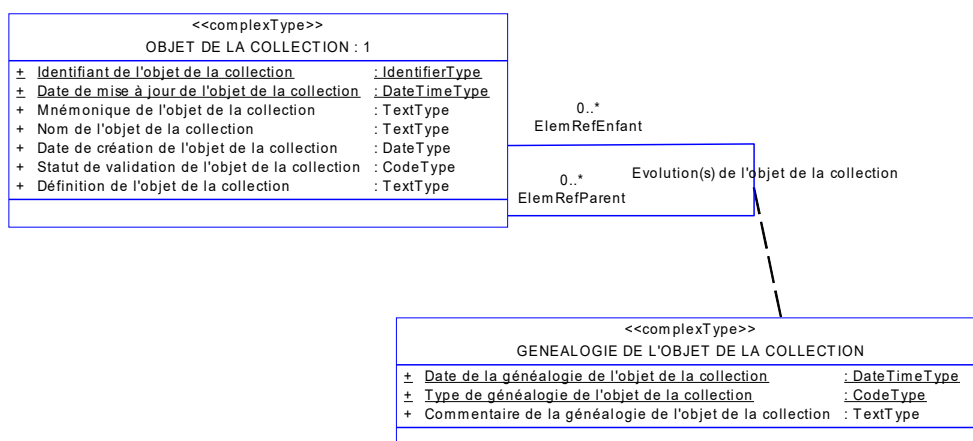


Image 0 : Concept d'évolution d'un élément de référentiel

L'expérience montre que l'attribut *Action* n'est pas toujours bien compris par les utilisateurs et producteurs de données du SIE. Cette note présente les difficultés rencontrées et propose une solution.

2. Scénarios d'échanges concernés

Un scénario d'échange est un ensemble d'indications techniques qui permet le dialogue entre des personnes et des systèmes informatiques. Un scénario d'échange établi par le Service d'administration nationale des données et référentiels sur l'eau (Sandre) est un document de spécification qui décrit les modalités d'échanges de données dans un contexte spécifique. Il sert à échanger des données en s'appuyant sur un format. Ce document détaille : la sémantique, le caractère obligatoire et facultatif, la syntaxe, des données échangées et les modalités techniques et organisationnelles de l'échange. Un scénario d'échange repose sur un ou plusieurs dictionnaires de données et se matérialise par des fichiers aux formats Xsd et Pdf.

2.1 Qualité des eaux superficielles continentales version 3

Par l'attribut *Action*, le créateur d'un fichier (i.e. jeu de données) conforme au scénario d'échange « [Qualité des eaux superficielles continentales version 3](#) » (cf. image n°1bis ci-dessous) peut indiquer les ajouts, les modifications et les suppressions éventuelles d'informations qu'il a réalisé dans son fichier depuis sa dernière diffusion mais seulement au niveau des balises Xml suivantes (cf. image n°1 ci-dessous) :

- station de mesure de la qualité des eaux de surface (cf. [StationMesureEauxSurface](#))
- opération de prélèvement physico-chimique et microbiologique (cf. [OperationPrel](#))
- opération de prélèvement biologique (cf. [OperationPrelBio](#))

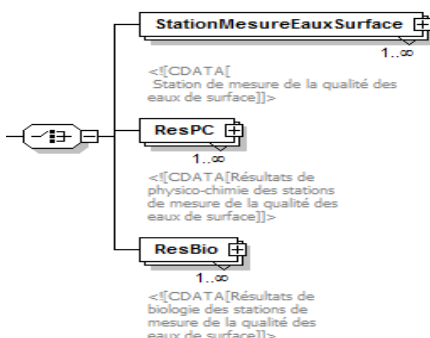


Image n°1: Extrait du schéma Xsd du scénario d'échange « [Qualité des eaux superficielles continentales version 3](#) » : http://xml.sandre.eaufrance.fr/scenario/quesu/3/sandre_sc_quesu.xsd

Un destinataire d'un fichier de données - respectant ce scénario d'échange - PEUT indiquer par la propriété « *Action* » présent au niveau des balises XML *StationMesureEauxSurface*, *OperationPrel* et *OperationPrelBio* que le lot de données (exemple un point d'eau et ses informations descriptives) a été **A** ajouté, **M** modifié voire **S** supprimé depuis sa dernière diffusion. Cet attribut prend pour valeurs possibles : **A**, **M** ou **S**.

Image n°1bis : Extrait du document Pdf du scénario d'échange [Qualité des eaux superficielles continentales version 3](#)

Par exemple, le jeu de données ci-dessous (cf. image n°2) comporte l'attribut *Action* qui est égal à « **A** » signifiant ajout. Mais, cela signifie-t-il :

- que l'opération de prélèvement a été ajoutée ?
- qu'un prélèvement de cette opération a été ajouté ?
- que des analyses ont été ajoutées à un prélèvement de cette opération ?
- etc.

```
<OperationPrel Action="A">
  <DateDebutOperationPrel>2012-01-01</DateDebutOperationPrel>
  <HeureDebutOperationPrel>10:00:00</HeureDebutOperationPrel>
  <DateFinOperationPrel>2012-01-23</DateFinOperationPrel>
  <HeureFinOperationPrel>12:00:00</HeureFinOperationPrel>
  <ComOperationPrel> </ComOperationPrel>
  <Prelevement> </Prelevement>
</OperationPrel>
```

Image n°2: Extrait du document Pdf du scénario d'échange [Qualité des eaux superficielles continentales version 3](#)

L'attribut *Action* de l'opération de prélèvement physico-chimique et microbiologique n'est applicable qu'à la balise XML mère [OperationPrel](#) (voir image n°3 ci-dessous). Il est donc impossible de connaître précisément quelles données, contenues dans les balises XML filles, ont évolué entre deux diffusions d'un fichier.

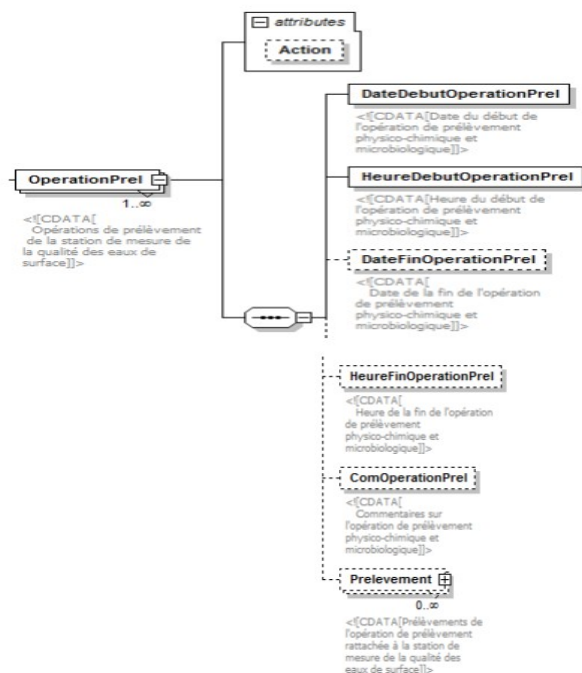


Image n°3: Extrait du schéma Xsd du scénario d'échange « [Qualité des eaux superficielles continentales version 3](http://xml.sandre.eaufrance.fr/scenario/quesu/3/sandre_sc_quesu.xsd) » :

2.2 Quantité et qualité des eaux souterraines version 2

Par l'attribut *Action*, le créateur d'un fichier (i.e. jeu de données) conforme au scénario d'échange « [Échanges de données quantité et qualité des eaux souterraines version 2](#) », (cf. image n°5 ci-dessous) peut indiquer les ajouts, les modifications et les suppressions éventuelles d'informations qu'il a réalisé dans son fichier depuis sa dernière diffusion mais seulement au niveau des balises XML (cf. image n°4 ci-dessous) :

- point d'eau (cf. [PointEau](#))
- prélèvement - qualitatif (cf. [Prelevement](#))
- chronique piézométrique - quantitatif (cf. [ChroniquePiezometrique](#))

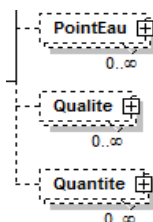


Image n°4 : Extrait du schéma Xsd du scénario d'échange « [Quantité et qualité des eaux souterraines version 2](http://xml.sandre.eaufrance.fr/scenario/qesout/2/sandre_sc_qesout.xsd) » :

La propriété « *attributes* » PEUT être ajoutée aux balises XML suivantes : PointEau, Prelevement , ChroniquePiezometrique. Elle permet de préciser - par l'attribut « Action » - que le lot de données (exemple un point d'eau et ses informations descriptives) a été **Ajouté**, **Modifié** voire **Supprimé** depuis sa dernière diffusion. Cet attribut prend pour valeurs possibles : **A**, **M** ou **S**.

Image n°5 : Extrait du scénario d'échange Sandre qualité des eaux souterraines Attribut Action:
<http://www.sandre.eaufrance.fr/urn.php?urn=urn:sandre:scenario-d-echanges:QESOUT:FRA::ressource:2.0:::pdf>

Par exemple, le jeu de données ci-dessous (cf. image n°6) comporte l'attribut *Action* qui est égal à « M » signifiant modification ; cela signifie-t-il :

- que les profondeurs du site de mesure ont changé ?
- que le résultat de la mesure a changé ?
- etc.

```
<ChroniquePiezometrique Action="M">
  <DateMesure>2011-01-06</DateMesure>
  <HeureMesure>09:00:00</HeureMesure>
  <ValeurMesure>11.00</ValeurMesure>
  <CdContinuiteMesure listID="72">2</CdContinuiteMesure>
  <ModeObtentionMesure listID="180">1</ModeObtentionMesure>
  <ConformiteAcqMesure listID="414">1</ConformiteAcqMesure>
  <StatutAcqMesure listID="415">2</StatutAcqMesure>
  <NatureMesure listID="604">N</NatureMesure>
  <Intervenant>
    <CdIntervenant schemeID="INT" schemeAgencyID="SIRET">17450988500029</CdIntervenant>
    <NomIntervenant>Direction Régionale de l'Environnement (D.I.R.E.N.) Centre</NomIntervenant>
  </Intervenant>
  <SiteMesurePiezo>
    <DebProfSiteMesurePiezo>23.87</DebProfSiteMesurePiezo>
    <FinProfSiteMesurePiezo>50.77</FinProfSiteMesurePiezo>
  </SiteMesurePiezo>
</ChroniquePiezometrique>
```

Image n°6 : Extrait du document Pdf du scénario d'échange [quantité et qualité des eaux souterraines version 2](#)

En effet, l'attribut *Action* de la chronique piézométrique n'est applicable qu'à la balise XML [ChroniquePiezometrique](#) (voir image n°7 ci-dessous). Il est donc impossible de connaître précisément quelles données, contenues dans les balises XML filles, ont évolué entre deux diffusions d'un fichier.

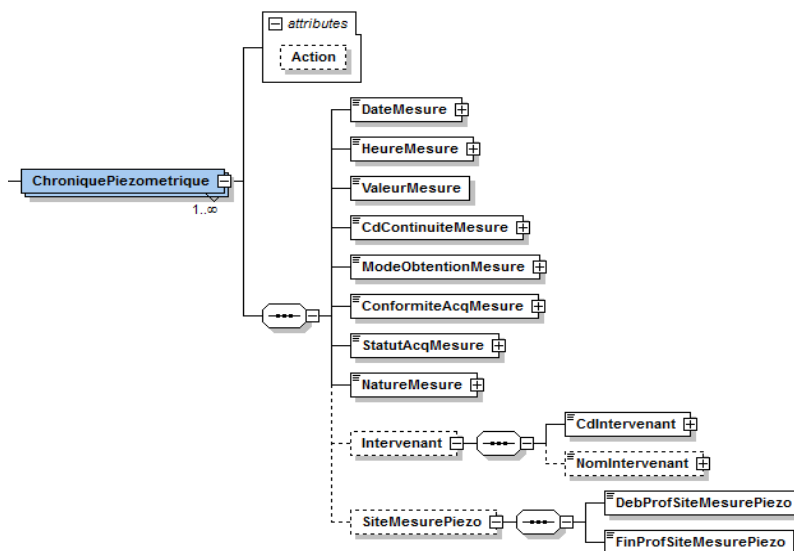


Image n°7 : Extrait du schéma Xsd du scénario d'échange «Quantité et qualité des eaux souterraines version 2» :
http://xml.sandre.eaufrance.fr/scenario/qesout/2/sandre_sc_qesout.xsd

3. Attributs date de création, de mise à jour des objets

Les objets de référence, comme les masses d'eau, disposent généralement d'attributs de traçabilité des évolutions (date de création, date de mise à jour et statut – cf. image 7bis) permettant de savoir quand un objet a été créé, mis à jour et « abandonné ». Ces informations pourraient être des doublons si elles coïncident avec les fréquences de diffusion du jeu de données - objet de l'attribut *Action*.

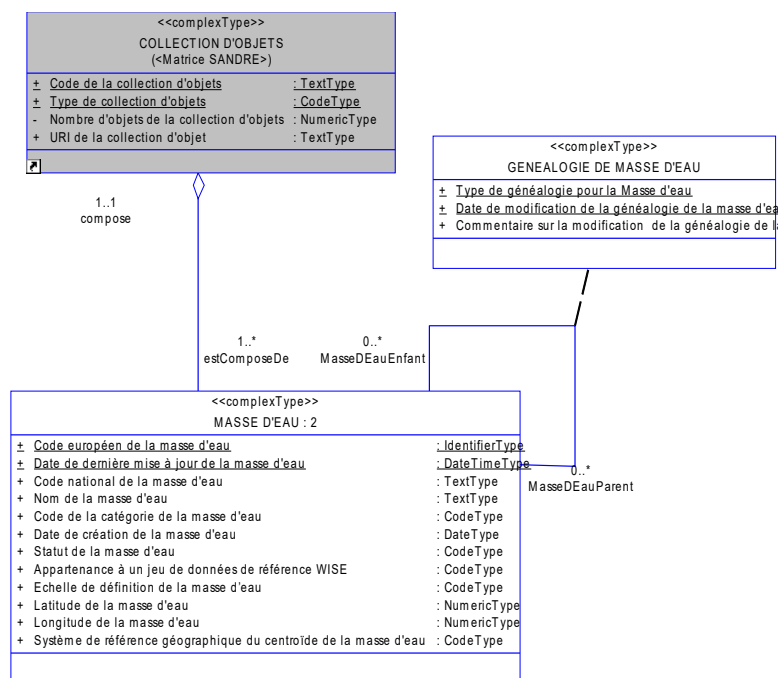


Image n°7bis : Extrait du dictionnaire de données masses d'eau version 1.3
<http://www.sandre.eaufrance.fr/urn.php?urn=urn:sandre:dictionnaire:MDO:FRA:::ressource:1.3:::pdf>

4. Scénario de service web concerné

Un scénario de service web est un ensemble d'indications techniques qui décrit l'interface (questions et réponses) d'un service web. Un scénario établi par le Service d'administration nationale des données et référentiels sur l'eau (Sandre) est un document de spécification qui décrit les modalités d'échanges de données d'un service web en s'appuyant sur un format. Il sert notamment à exposer des données stockées dans des bases de données par l'Internet. Un scénario technique repose sur un ou plusieurs scénarios d'échange et se matérialise par des fichiers aux formats WSDL et PDF.

Un producteur de données - qui implémente son service web selon le scénario de service web « [Qualité et quantité des eaux version 2.1](#) » - diffuse ses données (i.e. Jeux de données) selon les scénarios d'échange :

- Qualité des eaux superficielles continentales version 3
- Quantité et qualité des eaux souterraines version 2
- Qualité des eaux littorales version 1.1

L'utilisateur peut ainsi télécharger des jeux de données par le service web du producteur de données.

Plusieurs critères de requêtes sont possibles. Il est notamment possible de ne télécharger que les données qui ont été actualisées depuis une date fournie par l'utilisateur (cf. image n°8). L'utilisateur dispose alors d'un jeu de données actualisées à une date sans connaître précisément quelles données ont évolué. Pour cela, il devra se reporter aux scénarios d'échanges (cf. chapitres précédents).

dateMAJ	Date de mise a jour	Date de la dernière mise à jour réalisée par le consommateur de service.	0,1 (FACULTATIF)
---------	---------------------	--	------------------

Image n°8: Page 56 Scénario de service web qualité et quantité version 2.1 Opération GetData Paramètres de la requête : <http://www.sandre.eaufrance.fr/urn.php?urn=urn:sandre:scenario-de-service-web:monitoring:FRA::ressource:2.1:::pdf>

5. Perspectives

On pourrait combiner l'idée de la généalogie (cf. chapitre contexte) à celle de l'attribut *Action* et l'élargir. Dans cette hypothèse, il serait alors possible de savoir qu'un objet (exemple : un prélèvement) ou un de ses attributs (exemple : la date d'un prélèvement) a été ajouté/modifié/supprimé, par qui, pourquoi... Ces données seraient donc conservées dans le SIE. Le volume de données augmenterait pour des raisons de traçabilité et donc de qualité. Mais, la réalisation et l'entretien de ce dispositif seraient très certainement coûteux. Or, il n'est pas certain que la bancarisation et l'échange des données de traçabilité soient utiles pour tous les objets.

C'est pourtant le cas pour les objets de référence du SIE ; le [master data management](#) (MDM) du Sandre stocke les changements d'états des objets de référence (i.e. jeu de données de référence) jugés sensibles. Un jeu de données de référence est un ensemble de données de référence homogènes, codifiées et répertoriées. Les jeux de données de référence établis par le Sandre sont des données alphanumériques ou géographiques régulièrement évolutives. Les données alphanumériques sont des listes d'objets codifiés (par exemple, les codes des paramètres comme la demande en Chlore (1397), les méthodes, les intervenants, les nomenclatures, les stations de mesure...). Les données géographiques sont des objets géolocalisés (par exemple, les masses d'eau, les stations de mesure, les zones hydrographiques, les zonages réglementaires...). Chaque objet est associé à une liste d'informations complémentaires (code/identifiant, nom, localisation...).

Par conséquent, dans l'attente d'une réflexion plus globale des données de traçabilité du SIE, il est simplement proposé de créer une nouvelle nomenclature Sandre qui donnerait les définitions des valeurs possibles de l'attribut *Action*. Ces définitions seraient rédigées par le Secrétariat technique du Sandre et les producteurs qui ont implémenté cet attribut. Cette nomenclature serait ensuite validée par les Administrateurs Des Données (ADD). Concernant la généralisation de cet attribut au niveau des autres balises XML des scénarios d'échange, il est préférable d'en discuter au cas par cas.