

w.c.s. - Bug #78549

production d'xml invalide avec le switch à lxml

15 juin 2023 11:16 - Frédéric Péters

Statut:	Fermé	Début:	15 juin 2023
Priorité:	Haut	Echéance:	
Assigné à:	Pierre Ducroquet	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposé:	Non		

Description

<https://demarches-publik-famille.test.entrouvert.org/backoffice/workflows/68/>

exporté, donne :

Le statut de la demande que vous avez transmise vient de changer (il est passé de « {{ form_previous_status }} [ici caractère U0002]/body>

ce qui à l'import échoue logiquement,

```
File "/usr/lib/python3/dist-packages/wcs/api_export_import.py", line 286, in pre_install
    tree = ET.fromstring(element_content)
File "src/lxml/etree.pyx", line 3237, in lxml.etree.fromstring
File "src/lxml/parser.pxi", line 1896, in lxml.etree._parseMemoryDocument
File "src/lxml/parser.pxi", line 1784, in lxml.etree._parseDoc
File "src/lxml/parser.pxi", line 1141, in lxml.etree._BaseParser._parseDoc
File "src/lxml/parser.pxi", line 615, in lxml.etree._ParserContext._handleParseResultDoc
File "src/lxml/parser.pxi", line 725, in lxml.etree._handleParseResult
File "src/lxml/parser.pxi", line 654, in lxml.etree._raiseParseError
File "<string>", line 523
lxml.etree.XMLSyntaxError: PCDATA invalid Char value 2, line 523, column 59
```

Révisions associées

Révision fdd5601a - 15 juin 2023 21:39 - Pierre Ducroquet

global: replace our XML indent code with lxml (#78549)

Historique

#2 - 15 juin 2023 11:29 - Frédéric Péters

En local le fichier pickle d'origine fonctionne; et dans le changelog lxml, entre 4.6.3 (sur le serveur) et 4.9.2 (en local), je ne vois rien qui ressemble :/

#3 - 15 juin 2023 11:29 - Frédéric Péters

- Fichier 68 ajouté

#4 - 15 juin 2023 12:09 - Pierre Ducroquet

Frédéric Péters a écrit :

En local le fichier pickle d'origine fonctionne; et dans le changelog lxml, entre 4.6.3 (sur le serveur) et 4.9.2 (en local), je ne vois rien qui ressemble :/

Je pense qu'il faut plutôt regarder du côté de la libxml2 qui est utilisée par lxml.

Tu dois être en 2.9.14, le serveur en 2.9.10.

#5 - 15 juin 2023 13:10 - Pierre Ducroquet

Exporter uniquement le nœud problématique ne révèle aucun bug, il faut exporter tout le workflow pour que le problème se révèle.

#6 - 15 juin 2023 13:18 - Pierre Ducroquet

Si on n'appelle pas indent_xml, le bug ne se produit pas.

Étant donné qu'il y a une méthode `indent` dans `lxml`, et qu'elle n'a pas cet effet de bord (note : que je n'explique pas pour le moment), passer sur cette fonction est une solution rapide pour ce bug...

#7 - 15 juin 2023 13:28 - Pierre Ducroquet

Le document casse lorsqu'on modifie le "tail" du nœud `subject` qui précède le `body`.

Par contre là on part dans le mystique.

```
victim[2].tail = ""
b"\x02/body" in (ET.tostring(root)) ==> False
victim[2].tail = " "
b"\x02/body" in (ET.tostring(root)) ==> False
victim[2].tail = "  "
b"\x02/body" in (ET.tostring(root)) ==> False
victim[2].tail = "   "
b"\x02/body" in (ET.tostring(root)) ==> *True*
```

Mais si c'est `lxml` qui fait le changement, le `tail` ne pose pas de problème...

#8 - 15 juin 2023 13:42 - Robot Gitea

- Statut changé de *Nouveau* à *Solution proposée*

- Assigné à *mis* à Pierre Ducroquet

Pierre Ducroquet (pducroquet) a ouvert une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/wcs/pulls/385>
- Titre : replace our XML indent code with `lxml` (#78549)
- Modifications : <https://git.entrouvert.org/entrouvert/wcs/pulls/385/files>

#9 - 15 juin 2023 21:39 - Robot Gitea

- Statut changé de *Solution proposée* à *Solution validée*

Frédéric Péters (fpeters) a approuvé une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/wcs/pulls/385>

#10 - 15 juin 2023 21:39 - Robot Gitea

- Statut changé de *Solution validée* à *Résolu (à déployer)*

Frédéric Péters (fpeters) a mergé une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/wcs/pulls/385>
- Titre : replace our XML indent code with `lxml` (#78549)
- Modifications : <https://git.entrouvert.org/entrouvert/wcs/pulls/385/files>

#11 - 15 juin 2023 22:14 - Transition automatique

- Statut changé de *Résolu (à déployer)* à *Solution déployée*

#12 - 20 août 2023 04:42 - Transition automatique

Automatic expiration

Fichiers

68

19,3 ko

15 juin 2023

Frédéric Péters