

## w.c.s. - Development #4084

### Anonymisation des formulaires

10 décembre 2013 16:10 - Jérôme Schneider

<b>Statut:</b>	Fermé	<b>Début:</b>	10 décembre 2013
<b>Priorité:</b>	Normal	<b>Echéance:</b>	
<b>Assigné à:</b>	Thomas Noël	<b>% réalisé:</b>	0%
<b>Catégorie:</b>		<b>Temps estimé:</b>	0:00 heure
<b>Version cible:</b>		<b>Planning:</b>	
<b>Patch proposed:</b>			
<b>Description</b>			
Un premier jet pour l'anonymisation des formulaires dans l'admin. Ce patch enlève la fonction d'archivage et permet d'anonymiser les champs "libres", le user_id, le user_hash et les champs libres des évolutions.			

#### Révisions associées

##### Révision 08f6ab67 - 09 avril 2014 16:51 - Thomas Noël

admin: add 'anonymise' action (#4084)

#### Historique

##### #1 - 10 décembre 2013 17:01 - Thomas Noël

- Version cible mis à Au-quotidien 2014.5

Ca me parait la bonne façon de faire.

Je vois ça :

- nommer l'action "Anonymise handled forms" pour préciser que ça n'anonymise que les formulaires "terminés"
- y'a un "from wcs.fields import StringField" inutile dans formdata.py
- mais surtout il manque la partie SQL car on ajoute un attribut dans le formdata.

@Jérôme je peux me charger de ça et de la suite si tu es chargé par ailleurs, attribue-moi le ticket si tu veux.

##### #2 - 10 décembre 2013 17:11 - Frédéric Péters

Rapide commentaire perso, je préférerais la partie "suppression de la fonctionnalité d'archivage" dans un commit séparé.

##### #3 - 10 décembre 2013 17:31 - Frédéric Péters

Et faut plutôt faire le taf dans un afterjob (comme c'était le cas pour l'archivage).

##### #4 - 11 décembre 2013 11:59 - Jérôme Schneider

- Fichier 0001\_remove\_archive\_feature.patch ajouté

Voici un patch qui ne permet que de supprimer la fonction d'archivage.

##### #5 - 12 décembre 2013 09:29 - Jérôme Schneider

- Fichier 0002\_anonymise\_forms.patch ajouté

- Statut changé de Nouveau à En cours

Voici la seconde version qui tient compte des différentes remarques. Il faut appliquer le patch 0001\_anonymise\_forms.patch avant d'appliquer le 0002\_remove\_archive\_feature.patch.

##### #6 - 12 décembre 2013 11:33 - Frédéric Péters

Pour la valeur de "anonymised", plutôt que le time.localtime(), je prendrais le datetime.datetime.now() (ça nécessite derrière une modif du côté SQL).

##### #7 - 12 décembre 2013 13:52 - Thomas Noël

Moi j'aime bien que ça soit un timestamp en SQL (time.localtime()), comme pour receipt\_time et les dates des évolutions. J'AI RIEN DIT.

**#8 - 12 décembre 2013 15:50 - Jérôme Schneider**

- Fichier 0003\_anonymise\_forms.patch ajouté

La troisième version avec le passage de `time.localtime()` vers `datetime.datetime.now()`

**#9 - 18 décembre 2013 11:11 - Jérôme Schneider**

C'est OK pour que je pousse le patch 0001\_remove\_archive\_feature.patch puis le 0003\_anonymise\_forms.patch sur le dépôt ?

**#10 - 18 décembre 2013 16:54 - Frédéric Péters**

Tu parlais d'un truc de PostgreSQL de config de timezone ou un truc du même acabit, il en est quoi ?

**#11 - 19 décembre 2013 09:15 - Jérôme Schneider**

Je pensais que le timestamp avec gestion des timezone était obligatoire pour l'utilisation de `datetime` mais ce n'est pas le cas. J'ai choisi un timestamp sans timezone pour rester cohérent avec les autres timestamp de w.c.s. Maintenant si tu penses que ça peut être utile de rajouter la timezone je peux changer le type du champ anonymized.

**#12 - 19 décembre 2013 10:24 - Frédéric Péters**

Oui, je serais pour, ici, enregistrer l'information du fuseau horaire (et si je lis bien la doc de postgresql, il n'y a rien à faire, le `datetime.now()` retourne une valeur dans la timezone du serveur, et postgresql la convertira automatiquement pour stockage en utc).

**#13 - 14 janvier 2014 12:38 - Jérôme Schneider**

- Fichier 0002\_remove\_archive\_feature.patch ajouté

- Fichier 0004\_anonymise\_forms.patch ajouté

Je viens de mettre à jour les patches avec le master et de passer de `timestamp` à `timestampz`.

**#14 - 15 janvier 2014 12:20 - Frédéric Péters**

J'allais dire que c'était ok mais je me demande si la suppression de l'archivage peut se faire sans être compensée par l'ajout d'une fonctionnalité de suppression de données (d'un certain âge?) (et aussi une possibilité d'export des données ?).

Aussi l'anonymisation se fait sur tous les formulaires traités, il faudrait peut-être permettre de spécifier un âge minimum quand même. (?) Ou bien ce cas de figure se gère via une nouvelle action de workflow (déclenchée sur timeout, blabla). (?)

Bref, est-ce que pousser ça maintenant oblige à avoir dans la foulée des développements particuliers ?

**#15 - 15 janvier 2014 15:49 - Thomas Noël**

Frédéric Péters a écrit :

J'allais dire que c'était ok mais je me demande si la suppression de l'archivage peut se faire sans être compensée par l'ajout d'une fonctionnalité de suppression de données (d'un certain âge?) (et aussi une possibilité d'export des données ?).

Ouaip. Je dirais : on supprime l'archivage uniquement en mode SQL et seulement parce qu'il ne fonctionne pas encore. On le remet quand il remarchera...

Aussi l'anonymisation se fait sur tous les formulaires traités, il faudrait peut-être permettre de spécifier un âge minimum quand même. (?) Ou bien ce cas de figure se gère via une nouvelle action de workflow (déclenchée sur timeout, blabla). (?)

Idéalement selon moi:

- lors de l'anonymisation, permettre de choisir parmi les statuts terminaux (tous cochés par défaut)
- ajouter un champ "demande terminée depuis plus de "x" jours", avec x par défaut à 30 (on n'anonymise par défaut que les demandes dont le dernier traitement a eu lieu il y a plus de 30 jours).

**#16 - 04 février 2014 17:07 - Jérôme Schneider**

- Fichier 0003\_remove\_archive\_feature.patch ajouté

**#17 - 05 février 2014 15:11 - Jérôme Schneider**

- Fichier 0001\_anonymisation\_keep\_archiving.patch ajouté

Voici un nouveau patch qui prend en compte les différentes remarques à savoir garder la fonctionnalité d'archivage et la capacité de trier les données à anonymiser en fonction d'une date et des statuts terminaux.

## #18 - 12 février 2014 09:51 - Jérôme Schneider

Est ce que le dernier patch (0001\_anonymisation\_keep\_archiving.patch) vous semble OK pour être pousser dans le git ?

## #19 - 12 février 2014 10:10 - Thomas Noël

Jérôme Schneider a écrit :

Est ce que le dernier patch (0001\_anonymisation\_keep\_archiving.patch) vous semble OK pour être pousser dans le git ?

Tel quel non : il manque encore la gestion de l'affichage des données anonymisées dans les listings. A réfléchir un peu.

## #20 - 03 avril 2014 14:06 - Jérôme Schneider

- Fichier 0002\_anonymisation\_keep\_archiving.patch ajouté

Après discussion avec Thomas j'ai simplement ajouté un champ nommé "Anonymised" (non coché par défaut) dans les listings. Je joins le patch 0002\_anonymisation\_keep\_archiving.patch qui ajoute l'anonymisation complète et ce champ dans les listings.

## #21 - 06 avril 2014 10:43 - Frédéric Péters

```
+         r += htmltext('<li><a href="anonymise">%s</a></li>') % _('Anonymise handled forms')
```

Mineur : juste "Anonymise forms".

```
+     def anonymise(self):
+         all_forms = self.formdef.data_class().select()
+         for formdata in all_forms:
+             if not formdata.anonymised:
+                 formdata.anonymise()
+         if get_request().form.get('job'):
+             return self.anonymise_processing()
+     [...]
```

Majeur : c'est la méthode qui affiche le formulaire, c'est à mon avis une erreur (reste d'un ancien patch) d'avoir la procédure d'anonymisation en haut de celle-ci.

```
+         form.add(CheckboxesWidget, 'endpoints', title=_('Status of the requests to anonymise'),
```

Mineur : s/requests/forms/

```
+         def anonymise(self, job=None):
+             all_forms = self.formdef.data_class().select()
+             all_forms = [x for x in all_forms if x.status in self.status_ids]
+             all_forms = [x for x in all_forms if (
+                 x.evolution and x.evolution[-1].time < self.before_date) or (
+                 x.receipt_time < self.before_date)]
+             for formdata in all_forms:
+                 if not formdata.anonymised:
+                     formdata.anonymise()
```

Mineur : en terme de style, si le plan est de réduire et réduire la liste, le test "if not formdata.anonymised" devrait être faire de la même manière. En style alternatif, les différentes conditions peuvent être ajoutées dans la boucle, ainsi :

```
for formdata in self.formdef.data_class().select():
    if formdata.anonymised:
        continue
    if formdata.status not in self.status_ids:
        continue
    if (formdata.evolution and ...):
        continue
    if not formdata.anonymised:
        continue
    formdata.anonymise()
```

Comme tu veux.

En troisième voie il pourrait aussi y avoir l'utilisation de l'index sur les statuts, pour ne pas devoir passer par un select() général, mais ça demande plus de changements dans le code et comme on n'a pas d'exigence de rapidité ici, ce n'est pas nécessaire.

```
@@ -633,6 +635,7 @@ class FileField(WidgetField):
     key = 'file'
     description = N_('File Upload')
```

```
+ anonymise = False
+ widget_class = FileWithPreviewWidget
```

Majeur : les fichiers attachés doivent selon moi être anonymisés.

```
+ def anonymise(self):
+     for field in self.formdef.fields:
+         if field.anonymise:
+             self.data[field.id] = None
```

Majeur : il peut y avoir une variante `_display` (il faut tester `field.store_display_value`) et celle-ci devrait être anonymisée aussi (même si avec les champs existants aujourd'hui, ceux-là devraient tous avoir le `anonymise` à `False`).

Majeur : d'ailleurs, `ItemsField`, il devrait avoir son `anonymise` à `False`, non ?

```
+ r += htmltext('<td class="cell-anonymised">%s</td>') % anonymised
```

Mineur : je ferais remonter dans le `<tr>` un attribut `data-anonymised="true"` (ou une classe), qui pourrait par la suite servir à leur appliquer une apparence différente.

## #22 - 09 avril 2014 16:11 - Thomas Noël

- Fichier `0001-admin-add-anonymise-action.patch` ajouté

Voici une nouvelle version prenant en compte les remarques de Frédéric.

Notes :

Majeur : il peut y avoir une variante `_display` (il faut tester `field.store_display_value`) et celle-ci devrait être anonymisée aussi (même si avec les champs existants aujourd'hui, ceux-là devraient tous avoir le `anonymise` à `False`).

J'ai donc mis ça :

```
+ def anonymise(self):
+     for field in self.formdef.fields:
+         if field.anonymise:
+             self.data[field.id] = None
+             if field.store_display_value:
+                 self.data['%s_display' % field.id] = None
+ (...)
```

Majeur : d'ailleurs, `ItemsField`, il devrait avoir son `anonymise` à `False`, non ?

C'était bien le cas dans la proposition de Jérôme.

Mineur : je ferais remonter dans le `<tr>` un attribut `data-anonymised="true"` (ou une classe), qui pourrait par la suite servir à leur appliquer une apparence différente.

J'ai mis ça :

```
- r += htmltext('<tr class="status-%s-%s %s">') % (filled.formdef.workflow.id, filled.status, style)
+ )
+ data = ''
+ if filled.anonymised:
+     data += ' data-anonymised="true"'
+ r += htmltext('<tr class="status-%s-%s %s"%s>') % (filled.formdef.workflow.id,
+         filled.status, style, data)
```

## #23 - 09 avril 2014 16:12 - Thomas Noël

Thomas Noël a écrit :

Mineur : je ferais remonter dans le `<tr>` un attribut `data-anonymised="true"` (ou une classe), qui pourrait par la suite servir à leur appliquer une apparence différente.

... et je viens de me rendre compte que ma technique est idiote car elle donne un `data-anonymised="true"` ; ... partie à revoir.

## #24 - 09 avril 2014 16:20 - Frédéric Péters

Ouais, faut faire la substitution à l'intérieur du htmlentities() si tu veux faire ça ainsi...

**#25 - 09 avril 2014 16:23 - Thomas Noël**

- Fichier 0001-admin-add-anonymise-action.patch ajouté

Yep, voilà...

**#26 - 09 avril 2014 16:24 - Thomas Noël**

- Fichier 0001-admin-add-anonymise-action.patch supprimé

**#27 - 09 avril 2014 16:41 - Frédéric Péters**

data += ' data-anonymised="true"' ; j'aurais juste mis un =; mais c'est un détail, pour moi c'est ok ainsi.

**#28 - 09 avril 2014 16:53 - Thomas Noël**

- Statut changé de En cours à Solution déployée

- Assigné à changé de Jérôme Schneider à Thomas Noël

Poussé... à tester sur les machines de dev dès que possible.

(Note : il manque les traductions, as usual)

**#29 - 09 avril 2014 17:16 - Frédéric Péters**

En préparant les traductions j'ai noté qu'il y avait une chaîne "Requests ended before", j'ai changé ça pour utiliser "Forms" et non "Requests".

**#30 - 22 mai 2014 17:11 - Thomas Noël**

- Statut changé de Solution déployée à Fermé

**#31 - 26 août 2015 16:53 - Thomas Noël**

- Version cible Au-quotidien 2014.5 supprimé

**Fichiers**

---

0001_anonymise_forms.patch	10,5 ko	10 décembre 2013	Jérôme Schneider
0001_remove_archive_feature.patch	6,17 ko	11 décembre 2013	Jérôme Schneider
0002_anonymise_forms.patch	9,16 ko	12 décembre 2013	Jérôme Schneider
0003_anonymise_forms.patch	9,09 ko	12 décembre 2013	Jérôme Schneider
0002_remove_archive_feature.patch	6,39 ko	14 janvier 2014	Jérôme Schneider
0004_anonymise_forms.patch	9,09 ko	14 janvier 2014	Jérôme Schneider
0003_remove_archive_feature.patch	6,38 ko	04 février 2014	Jérôme Schneider
0001_anonymisation_keep_archiving.patch	10,1 ko	05 février 2014	Jérôme Schneider
0002_anonymisation_keep_archiving.patch	11,4 ko	03 avril 2014	Jérôme Schneider
0001-admin-add-anonymise-action.patch	12,2 ko	09 avril 2014	Thomas Noël