

Passerelle - Development #30010

connecteur API entreprise

22 janvier 2019 23:49 - Benjamin Dauvergne

Statut:	Fermé	Début:	22 janvier 2019
Priorité:	Normal	Echéance:	21 mars 2019
Assigné à:	Serghei Mihai	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Oui		
Description			
Voir https://doc.entreprise.api.gouv.fr/#entreprises			
Endpoints à traiter en priorité :			
Associations RNA Documents Association Extrait RCS Entreprise Etablissement			
Problème: pour les documents association, prévoir un proxy pour récupérer les documents par type mais les types ne semblent pas entièrement normés et certains documents n'en ont pas, il sera difficile de faire une API utilisable de façon simple depuis w.c.s.			
Demandes liées:			
Lié à Passerelle - Development #25432: Étendre les possibilités de documentat...		Nouveau	23 juillet 2018

Révisions associées

Révision 16100aa2 - 15 avril 2019 14:31 - Serghei Mihai

api_entreprise: add initial connector (#30010)

Historique

#3 - 23 janvier 2019 09:24 - Benjamin Dauvergne

- Tracker changé de Support à Development

#4 - 04 mars 2019 10:24 - Serghei Mihai

- Assigné à mis à Serghei Mihai

#5 - 14 mars 2019 17:51 - Serghei Mihai

- Echéance mis à 21 mars 2019

#6 - 20 mars 2019 12:05 - Serghei Mihai

- Patch proposed changé de Non à Oui

- Fichier 0001-api_entreprise-add-initial-connector-30010.patch ajouté

- Statut changé de Nouveau à Solution proposée

Je propose tous les documents qui ont un type défini.

#7 - 20 mars 2019 12:09 - Frédéric Péters

```
1. passerelle.apps.api_entreprise
```

Non, on ne fait pas ça.

```
from urlparse import urljoin
```

Sur les trucs évidents comme ceci, on essaie de penser à utiliser six.

```
li.connector.apientreprise a::before {
```

Non, la carte d'identité représente un usager, pas une asso ou entreprise.

#8 - 20 mars 2019 13:35 - Benjamin Dauvergne

Est-ce que pour les documents tu as pu voir si il y avait une régularité dans les types ? On pourrait en faire un référentiel à servir ? Si les nouveaux documents remplace les anciens ou si on se retrouve avec tout l'historique ?

#9 - 20 mars 2019 14:03 - Serghei Mihai

Des tests que j'ai faits je n'ai vu que les types que tu as listés plus haut.

Les nouveaux ne remplacent pas les anciens, une nouvelle ligne est rajoutée avec le timestamp du nouveau document.

On pourrait imaginer un référentiel des types documents pour pouvoir les filtrer par date (en isoformat).

#10 - 20 mars 2019 14:13 - Serghei Mihai

Déjà patch à jour avec les remarques de de Frédéric.

Pour l'icone j'ai choisi "fa-building-o" (<https://fontawesome.com/cheatsheet#regular>) qui me paraît la plus proche d'une entreprise.

#11 - 20 mars 2019 14:15 - Benjamin Dauvergne

Serghei Mihai a écrit :

Des tests que j'ai faits je n'ai vu que les types que tu as listés plus haut.

Les nouveaux ne remplacent pas les anciens, une nouvelle ligne est rajoutée avec le timestamp du nouveau document.

On pourrait imaginer un référentiel des types documents pour pouvoir les filtrer par date (en isoformat).

Idéalement dans une procédure on ne voudra que le dernier, mais à voir si en pratique on attache ça aux demandes ou on offre juste une fiche d'identité de l'association avec des liens vers tous les documents et un rendu de toutes les informations.

Niveau asso on s'intéresse à trois choses:

- les documents
- les membres du bureau
- d'éventuels coordonnées bancaires (je ne sais pas si ces APIs le porte)
Pour les entreprises je pense que seul le gérant intéresse les villes, mais peut-être le dernier bilan/CA.

#12 - 20 mars 2019 17:23 - Serghei Mihai

Benjamin Dauvergne a écrit :

Niveau asso on s'intéresse à trois choses:

- les membres du bureau

J'ai l'impression que c'est raté. L'API "associations RNA" (<https://doc.entreprise.api.gouv.fr/?json#associations-rna>) ne fourni qu'un dirigeant, et sur l'exemple de la doc mais aussi celle de mon asso de vélo (waldec: W751205868) les champs sont à null.

- d'éventuels coordonnées bancaires (je ne sais pas si ces APIs le porte)

Non.

Pour les entreprises je pense que seul le gérant intéresse les villes, mais peut-être le dernier bilan/CA.

Pour les entreprises on a les mandataires sociaux, rien sur le CA.

#13 - 20 mars 2019 17:28 - Benjamin Dauvergne

Serghei Mihai a écrit :

Benjamin Dauvergne a écrit :

Niveau asso on s'intéresse à trois choses:

- les membres du bureau

J'ai l'impression que c'est raté. L'API "associations RNA" (<https://doc.entreprise.api.gouv.fr/?json#associations-rna>) ne fournit qu'un dirigeant, et sur l'exemple de la doc mais aussi celle de mon asso de vélo (waldec: W751205868) les champs sont à null.

- d'éventuels coordonnées bancaires (je ne sais pas si ces APIs le porte)

Non.

Pour les entreprises je pense que seul le gérant intéresse les villes, mais peut-être le dernier bilan/CA.

Pour les entreprises on a les mandataires sociaux, rien sur le CA.

Yep je me doutais que les données étaient pourries, à mon avis ce serait plus intéressant d'avoir un bloc "association" dans toute démarche (et/ou un espace personne morale dans le portail agent) association où on affiche un maximum de choses, et le cas échéant la mention "manquante" pour certaines informations ; ça évitera les remontées des collectivités disant que ça ne marche pas ou que sais-je dans le cas où on choisirait plutôt de remplir des champs du formulaire avec ces informations.

#14 - 20 mars 2019 17:35 - Serghei Mihai

Tu veux dire le résultat de l'appel à <https://doc.entreprise.api.gouv.fr/?json#associations-rna> ?

#15 - 22 mars 2019 15:13 - Benjamin Dauvergne

Ces remarques globales étant faites, pour les dates ne faudrait-il pas normaliser pour des chaînes tous les timestamp (pour les identifier je dirai tous les clés qui commencent par "date_" et qui ont une valeur numérique, certaines finissent par _timestamp mais pas toute).

Pour les signatures je limiterai leur durée à quelque chose d'assez long, genre 1 semaine, on ne gêne pas le travail mais on ne republie pas non plus tout leur référentiel.

#16 - 22 mars 2019 15:56 - Serghei Mihai

- Statut changé de Solution proposée à En cours

Benjamin Dauvergne a écrit :

Ces remarques globales étant faites, pour les dates ne faudrait-il pas normaliser pour des chaînes tous les timestamp (pour les identifier je dirai tous les clés qui commencent par "date_" et qui ont une valeur numérique, certaines finissent par _timestamp mais pas toute).

Voici la liste des formats rencontrés:

```
"date_mise_a_jour": 1449183600,  
"date_reference": "2014",  
"date_creation_etablissement": 1108594800,  
"date_fermeture": 1315173600,  
"date_creation": 891381600,  
"date_naissance": "1965-01-27",  
"date_naissance": "",  
"date_naissance_timestamp": 0,  
"date_cessation": 1315173600,  
"date_immatriculation": "1998-03-27",  
"date_immatriculation_timestamp": 890953200,  
"date_extrait": "21 AVRIL 2017",  
"date": "2000-02-23",  
"date_timestamp": 951260400,  
"date_creation": "1993-02-11",  
"date_declaration": "2013-06-28",  
"date_publication": "1993-03-03",  
"mise_a_jour": "2013-06-28"
```

Le isoformat et le timestamp sont le plus communs. Donc à essayer isoformat et timestamp. Et si ça matche pas on retourne la chaîne d'origine.

Pour les signatures je limiterai leur durée à quelque chose d'assez long, genre 1 semaine, on ne gêne pas le travail mais on ne republie pas non plus tout leur référentiel.

Ok.

#17 - 22 mars 2019 16:24 - Benjamin Dauvergne

Serghei Mihai a écrit :

Benjamin Dauvergne a écrit :

Ces remarques globales étant faites, pour les dates ne faudrait-il pas normaliser pour des chaînes tous les timestamp (pour les identifier je dirai tous les clés qui commencent par "date_" et qui ont une valeur numérique, certaines finissent par _timestamp mais pas toute).

Voici la liste des formats rencontrés:

[...]

Juste convertir les timestamp en dates, timezone France, le reste bizarre on y touche pas.

#18 - 22 mars 2019 21:18 - Serghei Mihai

- Fichier *0001-api_entreprise-add-initial-connector-30010.patch* ajouté

- Statut changé de *En cours* à *Solution proposée*

Hop.

#19 - 25 mars 2019 10:34 - Benjamin Dauvergne

- context et object ne devraient pas être des paramètres globaux du connecteur mais surchargeables à chaque appel (pour pouvoir le réutiliser dans différents contextes et pour différents objets, demande de sub, réservations de salle, etc..), idem contexte sur une plate-forme multicollectivité on pourrait considérer d'y passer le nom de la collectivité
- Le code du endpoint document me paraît suspect:
 - il reste un print
 - on parcourt une liste on vérifie une signature sur un paramètre et je ne vois aucune interaction entre les deux.

http://git.entrouvert.org/passerelle.git/tree/passerelle/apps/api_entreprise/models.py?h=wip/30010-api-entreprise&id=10b7df32c3c4b5c79baf078e54b493a2ffb8fd7#n166

- Dans le endpoint documents_associations pareil, on signe juste timestamp et type (je peux comprendre que ce soit les seuls références qu'on ait) mais ensuite on utilise que timestamp comme id, on a un datetime naïf créé qui est juste mis dans le json (je ne sais pas pourquoi ça passe, peut-être a-t-on un encodeur JSON spécifique comme dans w.c.s., mais je n'en dépendrai pas ici).
 - Pour simplifier j'encoderais directement l'URL dans le dico signé donc juste `signing.dumps({'url': document['url']})` il n'est pas nécessaire d'encoder tout le reste, ce endpoint est juste un proxy (on peut encoder object et context par contre, le contexte/objet d'une récupération de document dépendra ainsi du contexte/objet au moment de la récupération de la liste des documents)
 - Le endpoint document ne doit pas être protégé par `can_access` il est publique, ainsi les URLs qu'on aura pourront être directement mise dans des interfaces (fiche association)
- Ce serait peut-être bien si possible de donner un exemple des contenus renvoyés pour ne pas avoir à croiser la doc de API entreprise avec celle du connecteur pour les utilisateurs (si on peut donner un exemple de retour dans la description du endpoint ce serait bien)

#20 - 25 mars 2019 10:37 - Benjamin Dauvergne

`normalize_dates` ignore les timestamp qu'on a justement sur le endpoint des documents (qui refait le job).

#21 - 29 mars 2019 15:05 - Serghei Mihai

- Fichier *0001-api_entreprise-add-initial-connector-30010.patch* ajouté

Benjamin Dauvergne a écrit :

`normalize_dates` ignore les timestamp qu'on a justement sur le endpoint des documents (qui refait le job).

J'ignorais le timestamp car je m'en servais comme id, mais ça peut être une date en effet.

Bonne idée pour utiliser url, object et context pour la récupération d'un fichier, merci.
Malheureusement dans endpoint on n'a rien pour donner des exemples de résultats des appels aux webservices.
Ça pourrait être objet d'un ticket à part.

#22 - 29 mars 2019 15:07 - Frédéric Péters

- Lié à *Development #25432: Étendre les possibilités de documentation des connecteurs* ajouté

#23 - 02 avril 2019 20:29 - Benjamin Dauvergne

- un print r qui traîne
- la validation d'url dans document me paraît inutile, on l'a signé c'est bon (et la signature on pourrait la valider hors de la boucle et s'arrêter là quand ça foire, enfin sans boucle le problème disparaît)
- en relisant la doc d'API entreprise je vois que contexte, objet et recipient sont obligatoires et doivent être pertinents alors je ne les rendrais pas du tout globaux finalement (surtout pas recipient qui doit être le SIRET/WALDEC du bénéficiaire) et je les mettrais explicitement dans la signature des appels avec des indications sur les valeurs attendues données dans la doc, ça devrait rendre le code plus clair et virer `get_context_and_object()` qui est un peu obscur.
 - pour contexte : APS pour aide publique simplifié, MPS pour marché publique simplifié, etc..
 - pour objet : le numéro de la demande dans w.c.s. ou d'un dossier
 - pour recipient: SIREN/SIRET/WALDEC du bénéficiaire
- j'ai toujours du mal avec le timestamp comme id, on a pas d'id faut faire avec, on utilisera jamais ce truc comme source donnée ça ne sert pas à grand chose, il faut juste renvoyer les données tels quels avec l'URL modifiée ; au mieux en imaginant un workflow où on aurait besoin d'un type de document particulier il faudrait un endpoint auquel on donne un type, un institution_id et qui renvoie le fichier le plus récent de ce type ou rien; parce que sinon j'ai du mal à imaginer un workflow où on viendrait utiliser les données renvoyées par document; on a un cahier des charges quelque part ? Je vois un ticket Arles mais qui ne précise pas les données qui les intéresse, si ça se trouve les documents ils s'en foutent.

Ce qui est clair c'est qu'au niveau document ça ne peut servir qu'à l'agent pas à la personne qui remplit le formulaire, à moins d'avoir un endpoint fonctionnant comme je le dis (pour un type et un id, ça renvoie le document le plus récent au format JSON/fichier attaché ou rien), mais en fait je ne sais pas trop si ça marche le prefill sur les champs fichiers.

#25 - 08 avril 2019 19:32 - Benjamin Dauvergne

Je serai pour ignorer la partie document et pousser tout de suite la partie qui a un intérêt immédiat pour Arles, donc juste ajouter les paramètres contexte, objet, recipient et ce sera ok.

#26 - 10 avril 2019 00:51 - Serghei Mihai

- Fichier `0001-api_entreprise-add-initial-connector-30010.patch` ajouté

Paramètres ajoutés.

#27 - 10 avril 2019 11:25 - Benjamin Dauvergne

```
# try to parse isoformatted dates
try:
    data[key] = datetime.fromtimestamp(int(data[key])).date()
except (ValueError, TypeError):
    pass
```

Le commentaire est pas cohérent avec le code qui suit, on parse un timestamp ici pas du format iso, en fait la date iso est parfois fourni avec sa version timestamp... et tu convertis aussi les timestamp des documents que tu vas utiliser comme id, je ne les toucherai pas.

Je ferai plutôt :

- si champ commence par "date" et ne finit pas par "timestamp" et contient un entier -> conversion vers "YYYY-mm-dd"
- si champ "timestamp" création d'un nouveau champ "datetime" avec conversion vers "YYYY-mm-dd HH:MM:ss" localtime (et donc pour les documents je retournerai aussi ce datetime)
- si timestamp 0, ne pas toucher (c'est une valeur particulière qui veut dire absent) ou alors convertir en None

#28 - 10 avril 2019 11:25 - Benjamin Dauvergne

- Statut changé de Solution proposée à En cours

#29 - 12 avril 2019 14:26 - Serghei Mihai

- Fichier `0001-api_entreprise-add-initial-connector-30010.patch` ajouté

- Statut changé de En cours à Solution proposée

Benjamin Dauvergne a écrit :

[...]

tu convertis aussi les timestamp des documents que tu vas utiliser comme id, je ne les toucherai pas.

A ce sujet tu dis plus haut: "j'ai toujours du mal avec le timestamp comme id, on a pas d'id faut faire avec, on utilisera jamais ce truc comme source donnée ça ne sert pas à grand chose" :) mais allez, il nous faut bien un id.

Je ferai plutôt :

- si champ commence par "date" et ne finit pas par "timestamp" et contient un entier -> conversion vers "YYYY-mm-dd"

Ok.

- si timestamp == 0, ne pas toucher (c'est une valeur particulière qui veut dire absent) ou alors convertir en None

Je préfère ne pas y toucher. C'est plus parlant que None.

#30 - 12 avril 2019 14:59 - Benjamin Dauvergne

- Ça ne va pas marcher avec "date_ref": "2014" :

```
if isinstance(data[key], int) or data[key].isdigit():
    try:
        data[key] = datetime.fromtimestamp(int(data[key])).date()
    except (ValueError, TypeError):
        pass
```

ça va faire un .fromtimestamp(2014) qui n'est pas du tout ce qu'on souhaite, j'en resterai à "c'est un entier" et c'est tout

- Ça on aimerait conserver le préfixe de la clé je pense :

```
if key.endswith('timestamp'):
    # timestamps can be integers or strings
    # convert only if positive values
    if int(data[key]) > 0:
        try:
            object_datetime = make_aware(datetime.fromtimestamp(int(data[key])))
        except (ValueError, TypeError):
            pass
    if object_datetime is not None:
        data['datetime'] = object_datetime
```

et donc plutôt

```
else:
    data[key[:-len('timestamp')] + 'datetime'] = object_datetime
```

en virant la suite.

#31 - 12 avril 2019 15:03 - Benjamin Dauvergne

- Statut changé de Solution proposée à En cours

#32 - 15 avril 2019 09:55 - Serghei Mihai

- Fichier 0001-api_entreprise-add-initial-connector-30010.patch ajouté

- Statut changé de En cours à Solution proposée

J'ai du rêver qu'il y avait des timestamps sous forme de chaîne.
Bonne idée de garder de prefixer le datetime par le nom du champ.

#33 - 15 avril 2019 14:12 - Benjamin Dauvergne

Une dernière chose, tu ne retournes pas le datetime, des documents, qu'on vient de créer; il faudrait :

```
+ item['datetime'] = item['datetime']
```

#34 - 15 avril 2019 14:23 - Serghei Mihai

Si. Dans les tests:

```
for document in data:
    assert 'id' in document
    assert 'text' in document
    assert 'url' in document
    assert 'type' in document
    assert 'datetime' in document
```

#35 - 15 avril 2019 14:27 - Benjamin Dauvergne

- Statut changé de Solution proposée à Solution validée

Au temps pour moi. Ack.

#36 - 15 avril 2019 14:32 - Serghei Mihai

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

```
commit 16100aa2fb5aa18f01231f6ce361bcbc91ccb3b1 (origin/master, origin/HEAD)
Author: Serghei Mihai <smihai@entrouvert.com>
Date: Tue Mar 19 15:37:04 2019 +0100
```

```
api_entreprise: add initial connector (#30010)
```

#37 - 23 avril 2019 11:48 - Serghei Mihai

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

Fichiers

0001-api_entreprise-add-initial-connector-30010.patch	26,5 ko	20 mars 2019	Serghei Mihai
0001-api_entreprise-add-initial-connector-30010.patch	27,8 ko	22 mars 2019	Serghei Mihai
0001-api_entreprise-add-initial-connector-30010.patch	29,3 ko	29 mars 2019	Serghei Mihai
0001-api_entreprise-add-initial-connector-30010.patch	31,4 ko	09 avril 2019	Serghei Mihai
0001-api_entreprise-add-initial-connector-30010.patch	32,3 ko	12 avril 2019	Serghei Mihai
0001-api_entreprise-add-initial-connector-30010.patch	32,3 ko	15 avril 2019	Serghei Mihai