



## Documentation technique Stock GetContacts Anonymized

---

### 1. Présentation

Ce document décrit le protocole de mise à disposition du stock de données de contacts anonymisés societeinfo. Ce flux permet l'intégration des données de contact societeinfo dans un système tiers à des fins de recherche basées sur le type de poste ou d'extraction massive.

Ce flux est uniquement accessible aux partenaires societeinfo. Un compte partenaire et une clé d'API spécifique est nécessaire pour extraire de l'API societeinfo les données complètes non-anonymisées.

Le jeu de donnée inclue les informations de contacts suivantes :

- Le SIREN de la société
- L'identifiant du contact
- Une clé de hash pour détecter les mises à jour du contact
- Flag d'existence Civilité, Prénom, NOM
- Poste
  - Libellé du rôle linkedin ET/OU libellé du mandat
  - Domaine
  - Niveau hiérarchique
- Flag d'existence de date de naissance
- Linkedin
  - Flag d'existence d'url du profil
  - Date d'extraction
- Email
  - Flag d'existence d'un email
  - Score de pertinence
  - Date de test de l'email

- Résultat du test de l'email
- Flag Webmail

Les algorithmes societeinfo parcourent quotidiennement le web pour extraire ces informations, un flux de mise à jour quotidien est à disposition pour permettre l'actualisation des données.

## **2. Format de donnée**

Le flux GetContact Anonymized est mise à disposition sous la forme d'un fichier csv compressé au format gzip.

La donnée est également disponible sous forme JSON sur l'API societeinfo en positionnant le paramètre « show\_all\_anonymized »

<https://societeinfo.com/api-doc/#get-contacts-by-company>

Colonne	Description	Valeurs possibles	Optionnel
<b>siren</b>	SIREN de la société		NON
<b>contact_type</b>	Type de contact	Personal	NON
<b>contact_source</b>	Source du contact	legal, linkedin, domain_registrant	NON
<b>id</b>	Identifiant unique du contact		NON
<b>hashcode</b>	Clé de hash de la donnée de contact		NON
<b>role</b>	Mandat ET/OU Rôle linkedin		NON
<b>civility</b>	Civilité	Monsieur/Madame	OUI
<b>first_name</b>	Flag existence Prénom	Y/N	NON
<b>last_name</b>	Flag existence Nom	Y/N	NON
<b>birthdate</b>	Flag existence Date de naissance	Y/N	OUI
<b>linkedin_url</b>	Flag existence Url Profil linkedin	Y/N	OUI
<b>linkedin_extraction_date</b>	Date extraction linkedin		OUI
<b>email</b>	Flag existence Email	Y/N	OUI
<b>email_score</b>	Score de pertinence de l'email	Double entre 0.0 et 1.0	OUI
<b>email_test_result</b>	Résultat du test email	Safe to Send, Risky, Invalid, Unknown	OUI
<b>email_test_reason</b>	Raison du test email	Deliverable, Bounce, Accept All, Unreachable	OUI
<b>email_test_date</b>	Date du test email		OUI
<b>email_webmail</b>	Indique si l'email est un webmail (google.fr, yahoo.fr...)		OUI
<b>phone</b>	Flag existence téléphone (Non disponible pour des raisons RGPD)	N	OUI
<b>role_domain_code</b>	Code domaine d'activité	Voir 6.1 – Domaine d'activité	OUI
<b>role_level_code</b>	Code Niveau/séniorité	Voir 6.2 – Niveau hiérarchique	OUI

### **3. Couverture du jeu de donnée**

L'extraction est limitée aux sièges des entreprises et aux sociétés actives.

Le jeu de donnée n'inclue que les contacts avec un profil linkedin ET/OU un email.

#### **3.1 Stock**

Le fichier stock est produit le premier samedi de chaque mois, il est disponible sous le chemin suivant :

`contact_anonymized/stock/contact_anonymized_stock_AAAAMMJJ.csv.gz`

Il est recommandé de purger l'ensemble des contacts à chaque mise à jour de stock.

#### **3.2 Update**

Un fichiers de mise à jour avec la liste des contacts actifs par SIREN est produit tous les jours, il est disponible sous le chemin suivant :

`contact_anonymized/update/contact_anonymized_update_AAAAMMJJ.csv.gz`

Un fichiers de mise à jour avec la liste des contacts obsolètes par SIREN est produit tous les jours, il est disponible sous le chemin suivant :

`contact_anonymized/update/contact_anonymized_dead_update_AAAAMMJJ.csv.gz`

Il est indispensable de synchroniser au quotidien son référentiel avec le flux d'update car l'API ne permettra pas la restitution d'un contact obsolète.

## 4. Intégration avec la donnée non-anonymisée

La donnée de contact non anonymisée s'extrait par API en requêtant avec les paramètres SIREN et CONTACT\_ID

<https://societeinfo.com/api-doc/#get-contacts-by-company>

Exemple :

[https://societeinfo.com/app/rest/api/v2/contacts.json/SIREN?contact\\_ids=ID1,ID2](https://societeinfo.com/app/rest/api/v2/contacts.json/SIREN?contact_ids=ID1,ID2)

## 5. Classification des postes

### 5.1 Domaine d'activité

Code	Libellé	Disponible dans l'IHM
DG	Direction Générale	Oui
ADMIN_FINANCE	Administration et Finance	Oui
ACHAT	Achats	Oui
RH	Ressources Humaines	Oui
IT	Informatique	Oui
MARKETING	Marketing	Oui
COMMUNICATION	Communication	Oui
CONFORMITE	Conformité	Oui
COMMERCE	Commerce	Oui
JURIDIQUE	Juridique	Oui
R_D	Recherche et Développement	Oui
SAV	Relation et support client	Oui
PROD_LOGISTIQUE	Production et Logistique	Oui
AGRI	Agricole	Non
ARMEE	Armée et Défense	Non
ART	Art et Design	Non
AUTOMOBILE	Automobile	Non
BANQ_ASSUR	Banque Assurance	Non
BEAUTE	Soin Beauté Cosmétique	Non
BTP	Construction	Non

CAISSIER	Personnel de caisse	Non
CHR	Café Hotellerie Restauration	Non
FORMATION_ENSEIGN	Formation et Enseignement	Non
FUNERAIRE	Services funéraires	Non
LOISIR	Loisir	Non
MAGASIN	Magasin	Non
PHARMA	Pharmacie	Non
PRESSING	Pressing	Non
PUBLIC_SOCIAL	Administration publiques et sociales	Non
SECURITE	Sécurité	Non
SANTE	Santé et Médical	Non
TEXTILE	Textile	Non

## 5.2 Niveau hiérarchique

Code	Libellé
MANDATAIRE	Mandataire
DIRIGEANT	Dirigeant
GERANT	Gérant
DIRECTEUR	Directeur
RESPONSABLE	Responsable
ASSIST	Assistant
COLLAB	Collaborateur
AUTRE	Autre

## 6. Exemple de donnée

```
siren;contact_type;contact_source;id;hascode;role;first_name;last_name;birthdate;linkedin_u
rl;linkedin_extraction_date;email;email_score;email_test_result;email_test_reason;email_tes
t_date;phone;civility
005780044;Personal;legal|domain_registrant;0057800441227373930;1291968454;Président | Webma
ster;Y;Y;N;N;Y;0 89;Safe to Send;Deliverable;2020-04-10;Y
005780044;Personal;legal|linkedin;0057800442012121180;116529871;Directeur général | Directe
ur;Y;Y;N;Y;2020-04-15;N;;;;;Y
005780044;Personal;linkedin;005780044562173561;257795606;Chef d'équipe;Y;Y;N;Y;2020-04-15;N
;;;;;Y
005780044;Personal;linkedin;0057800441935288371;260285470;Responsable administration commer
ciale;Y;Y;N;Y;2020-04-15;N;;;;;Y
007050149;Personal;legal|linkedin;007050149268764966;35009762;Gérant | Associé gérant;Y;Y;Y
;Y;2020-04-12;N;;;;;Y
```

```

007050149;Personal;legal|linkedin;007050149671671332;1294929989;gerant;Y;Y;N;Y;2020-04-12;N
;;;;;Y
007280217;Personal;linkedin;007280217204676714;22900977;Aide a la personne;Y;Y;N;Y;2020-04-
15;N;;;;;Y
006480206;Personal;linkedin;006480206579645411;1834573904;Consultant;Y;Y;N;Y;2019-10-01;N;
;;;;;Y
006480206;Personal;linkedin;00648020618504899;529614423;Compte clé gms & responsable des ve
ntes;Y;Y;N;Y;2020-04-16;N;;;;;Y
006480206;Personal;linkedin;006480206715965011;1665993316;Responsable administratif et fina
ncier;Y;Y;N;Y;2019-10-01;N;;;;;Y
006480206;Personal;linkedin;006480206987176299;1561381914;Responsable de production et extr
action;Y;Y;N;Y;2020-04-16;N;;;;;Y
006180301;Personal;linkedin;0061803011031799553;474174447;Responsable achats;Y;Y;N;Y;2019-1
0-01;Y;0 70;Risky;Accept All;2020-04-16;;Y
006180301;Personal;legal|linkedin;0061803011894153170;2078290481;Directeur général | Presid
ent du directoire | Société financière atlantic;Y;Y;N;Y;2020-04-16;Y;0 16;Risky;Accept All;
2020-01-07;;Y
006180301;Personal;linkedin;006180301836232932;673060250;Responsable achats;Y;Y;N;Y;2020-04
-16;Y;0 16;Risky;Accept All;2020-01-07;;Y

```

## 7. Métadonnée

Des fichiers de métadonnée sont générés indiquant les dates de mise à jour des différentes sources légales ainsi que des informations techniques sur l'archive.

```

{
  "id": "societeinfo/StockGetContactsAnonymized",
  "title": "Societeinfo Stock GetContacts Anonymized",
  "description": "Dataset including anonymized employees and manager profiles for all french
companies",
  "scope": "Stock",
  "creation_date": "2021-10-02T07:30:09",
  "path": "/data/repository/ftp/contact_anonymized/stock/contact_anonymized_stock_20211002.
csv",
  "compression": "gz",
  "resources": [
    {
      "file_name": "contact_anonymized_stock_20211002.csv",
      "file_location": "/data/repository/ftp/contact_anonymized/stock/",
      "file_format": "csv",
      "file_size": "1037729224"
    }
  ],
  "external_sources": [
    {
      "dataset_id": "inpi/imr/tc",
      "author": "INPI",
      "license": "licence-ouverte-open-licence v2.0",
      "update_status": "COMPLETED",
      "update_date": "2021-10-02T06:27:13",
      "as_of_date": "2021-10-01"
    },
    {
      "dataset_id": "inpi/bilans",
      "author": "INPI",
      "license": "licence-ouverte-open-licence v2.0",

```

```
    "update_status": "COMPLETED",
    "update_date": "2021-10-02T06:56:18",
    "as_of_date": "2021-10-01"
  },
  {
    "dataset_id": "insee/sirene",
    "author": "INSEE",
    "license": "licence-ouverte-open-licence v2.0",
    "update_status": "COMPLETED",
    "update_date": "2021-10-02T01:22:02",
    "as_of_date": "2021-09-30"
  },
  {
    "dataset_id": "bodacc",
    "author": "Direction de l'information légale et administrative",
    "license": "licence-ouverte-open-licence",
    "update_status": "COMPLETED",
    "update_date": "2021-10-02T06:47:04",
    "as_of_date": "2021-09-30"
  }
]
}
```