

Documentation technique pour les données issues de BOFiP

Objectif du document : Ce document contient une documentation technique générique concernant les fichiers mis à disposition dans le cadre des licences de réutilisation des données provenant de BOFiP.

<i>Version / Date</i>	<i>Émetteur</i>	<i>Suivi des modifications</i>
1.0 / 06.12.2019	DGFIP	Initialisation du document
1.1 / 15.06.2020	DGFIP	Ajustements suite évolutions techniques

Table des matières

1 PRÉAMBULE.....	3
2 FICHIERS FOURNIS.....	3
3 ORGANISATION DES RÉPERTOIRES.....	4
4 STRUCTURE DU FICHIER DOCUMENT.XML.....	5
4.1 DESCRIPTION DU FICHIER SOURCE (PART).....	5
4.2 MÉTA-DONNÉES DUBLIN CORE (DC:DUBLINCORE).....	5
4.3 MÉTA-DONNÉES BOFiP (BOFiP:BODGFIP).....	6
4.4 EXEMPLE DE FICHIER DOCUMENT.XML.....	7
5 STRUCTURE DU DOCUMENT SOURCE.....	8

1 PRÉAMBULE

Le site *BOFiP* a pour but d'offrir un accès public à la doctrine fiscale. Dans le cadre de l'Open Data, il est prévu de mettre à disposition les documents constituant cette doctrine fiscale. Ainsi un fichier archive de type tar.gz est constitué afin de regrouper les documents doctrinaux ainsi que les méta-données qui les accompagnent.

2 FICHIERS FOURNIS

Un fichier stock contenant l'intégralité des données du référentiel BOFiP (dans leur version en cours) est extrait une fois par mois. Celui-ci prend un nom tel que défini ci-dessous :

`bofip_stock_live_AAAAMMJJ`

avec :

AAAAMMJJ correspondant à la date de constitution du fichier.

Un fichier flux est extrait toutes les semaines et contient l'intégralité des données du référentiel BOFiP émises depuis le dernier fichier flux (dans leur version en cours). Celui-ci prend un nom tel que défini ci-dessous :

`bofip_flux_live_AAAAMMJJ_BBBBNKK`

avec :

AAAAMMJJ correspondant à la date de début de la période d'extraction,

BBBNKK correspondant à la date de fin de la période d'extraction.

Tous les fichiers archives ainsi constitués sont accompagnés d'un fichier de contrôle MD5 permettant de valider le fichier archive source.

3 ORGANISATION DES RÉPERTOIRES

L'export des données produit deux fichiers pour chaque document du référentiel BOFiP. Ceux-ci sont hébergés au sein d'une arborescence de fichiers.

La structure de stockage ainsi définie est significative en ce sens qu'une répartition est effectuée en tenant compte du type de document, ainsi que de sa catégorie (Commentaire, Barème, ...) et de sa série (BA, IR, ...) lorsqu'il s'agit d'un contenu doctrinal. De plus, les différentes versions sont stockées en fonction de leur date de publication.

Ce mode de répartition est enregistré dans la structure suivante :

BOFiP	Répertoire racine
<ul style="list-style-type: none"> └ documents <ul style="list-style-type: none"> └ Fichier <ul style="list-style-type: none"> └ identifiant document <ul style="list-style-type: none"> └ aaaa-mm-jj <ul style="list-style-type: none"> └ document.xml └ data.xxx └ Contenu <ul style="list-style-type: none"> └ type document <ul style="list-style-type: none"> └ serie <ul style="list-style-type: none"> └ identifiant document <ul style="list-style-type: none"> └ aaaa-mm-jj <ul style="list-style-type: none"> └ document.xml └ data.html └ Image <ul style="list-style-type: none"> └ identifiant document <ul style="list-style-type: none"> └ aaaa-mm-jj <ul style="list-style-type: none"> └ document.xml └ data.xxx └ PlanClassement <ul style="list-style-type: none"> └ identifiant document <ul style="list-style-type: none"> └ aaaa-mm-jj <ul style="list-style-type: none"> └ document.xml └ data.xml └ infos <ul style="list-style-type: none"> └ data.dtd └ document.dtd └ plan.dtd 	<ul style="list-style-type: none"> Documents sources <ul style="list-style-type: none"> Documents attachés <ul style="list-style-type: none"> Identifiant du document Date de publication Méta-données Document attaché Documents contenus doctrinaux <ul style="list-style-type: none"> Type de document Série <ul style="list-style-type: none"> Identifiant du document Date de publication Méta-données Document contenu Documents images <ul style="list-style-type: none"> Identifiant du document Date de publication Méta-données Document image Document plan de classement <ul style="list-style-type: none"> Identifiant du document Date de publication Méta-données Document plan de classement Documents techniques <ul style="list-style-type: none"> dtd du document contenu dtd du document des méta données dtd du document plan de classement

La partie document stocke tous les documents produits tandis que la partie infos héberge les dtd relatives aux fichiers de la doctrine fiscale, du document d'accompagnement et du plan de classement.

4 STRUCTURE DU FICHIER DOCUMENT.XML

Le fichier « *document.xml* », qui contient les méta-données d'accompagnement du document source, est au format xml et possède un formalisme dans lequel deux espaces de noms distincts cohabitent :

- l'espace de nom Dublin Core avec, a minima, un certain nombre de rubriques standardisées,
- l'espace de nom BOFiP avec quelques méta-données propres à l'application.

Ce fichier contient trois grandes parties qui permettent de gérer le fichier source qu'il accompagne ;

- la description du fichier source,
- la description des méta-données Dublin Core,
- la description des méta-données BOFiP.

4.1 Description du fichier source (parts)

Le paragraphe <parts> décrit le type et la référence du document source :

- Type / mime du document source: **Balise** : <dataType>
- Référence du document source; **Balise** : <dataRef>

4.2 Méta-données Dublin Core (dc:dublincore)

Le paragraphe <dc:dublincore> contient quatorze des quinze méta-données préconisées par le standard Dublin Core :

- Identifiant : **Balise** : <dc:identifiant>
- Titre du document : **Balise** : <dc:title>
- Bureaux producteurs : **Balise** : <dc:creator>
- Éditeur : **Balise** : <dc:publisher>
- Langue : **Balise** : <dc:language>
- Format : **Balise** : <dc:format>
- Source : **Balise** : <dc:source>
- Condition d'usage et de réutilisation : **Balise** : <dc:rights>
- Couverture du contenu : **Balise** : <dc:coverage>
- Code de classification : **Balise** : <dc:subject>
- Date de publication : **Balise** : <dc:date>
- Lien entre documents : **Balise** : <dc:relation>

à l'exception de :

- Catégorie de document : **Balise** : <dc:type>

Toutes les méta-données ci-dessus ne sont pas obligatoirement servies.
Certaines d'entre elles peuvent être multivaluées.

Les liens entre documents sont structurés de la manière suivante :

<Nature du document>.<Type de document>:<identifiant du document>

Ces liens sont caractérisés par un attribut *type* dans la balise <dc:relation> qui précise la nature de la relation entre les documents :

- references le document courant référence le document du lien,
- isReferencedBy le document courant est référencé par le document du lien,
- requires le document courant requiert le document du lien,
- isRequiredBy le document courant est requis par le document du lien.

4.3 Méta-données BOFiP (*bofip:bodgfip*)

Le paragraphe `<bofip:bodgfip>` contient la liste des méta-données spécifiques issues de l'application BOFiP :

Description	Balise
Type de document contenu	<code><bofip:contenu_type></code>
Niveau du document	<code><bofip:contenu_niveau></code>
Identifiant d'un document contenu	<code><bofip:contenu_id></code>
Directeur de publication	<code><bofip:contenu_directeur_publication></code>
Code ISBN BOFiP	<code><bofip:contenu_isbn></code>

Les types de document contenu répertoriés sont les suivants :

- Commentaire
- Barème
- Formulaire
- Lettre type; modèle
- Cartographie
- Autres annexes

4.4 Exemple de fichier document.xml

Un exemple de fichier « *document.xml* » est fourni ci-dessous :

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<document type="Contenu">
  <parts>
    <part>
      <dataType>text/html</dataType>
      <dataRef>data.html</dataRef>
    </part>
  </parts>
  <dc:dublincore xmlns:dc="http://purl.org/dc/elements/1.1">
    <dc:title>IF - Cotisation foncière des entreprises - Champ d'application -
    Personnes et activités exonérées - Exonérations facultatives temporaires
    accordées dans le cadre de la politique de la ville - Activités commerciales dans
    les quartiers prioritaires de la politique de la ville</dc:title>
    <dc:date>2018-07-04</dc:date>
    <dc:creator>DGFIP</dc:creator>
    <dc:subject>IF</dc:subject>
    <dc:subject>Cotisation foncière entreprises</dc:subject>
    <dc:publisher>Ministère du budget - DGFIP (France)</dc:publisher>
    <dc:format>text/xml</dc:format>
    <dc:identifiant>10166-PGP</dc:identifiant>
    <dc:identifiant>http://bofip.impots.gouv.fr/bofip/10166-PGP.html?
    identifiant=B0I-IF-CFE-10-30-50-60-20180704</dc:identifiant>
    <dc:source>Bulletin officiel des finances publiques - Impôts</dc:source>
    <dc:language>fr-FR</dc:language>
    <dc:relation type="references">Contenu.Autres annexes:1077-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:2671-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:389-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:4116-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:4320-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:4403-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:4490-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:7797-PGP</dc:relation>
    <dc:relation type="references">Contenu.Commentaire:9617-PGP</dc:relation>
    <dc:relation
    type="isReferencedBy">Contenu.Commentaire:10167-PGP</dc:relation>
    <dc:relation type="isReferencedBy">Contenu.Commentaire:4320-PGP</dc:relation>
    <dc:relation type="isReferencedBy">Contenu.Commentaire:4490-PGP</dc:relation>
    <dc:relation type="isReferencedBy">Contenu.Commentaire:5903-PGP</dc:relation>
    <dc:coverage>République française</dc:coverage>
    <dc:rights>Accès libre - réutilisation soumise à licence</dc:rights>
  </dc:dublincore>
  <bofip:bodgfip xmlns:bofip="http://bofip.impots.gouv.fr">
    <bofip:contenu_type>Commentaire</bofip:contenu_type>
    <bofip:contenu_niveau>Enfant</bofip:contenu_niveau>
    <bofip:contenu_id>B0I-IF-CFE-10-30-50-60</bofip:contenu_id>
    <bofip:contenu_directeur_publication>Bruno Parent, directeur général des
    finances publiques</bofip:contenu_directeur_publication>
    <bofip:contenu_isbn>2262-1954</bofip:contenu_isbn>
  </bofip:bodgfip>
</document>
```

5 STRUCTURE DU DOCUMENT SOURCE

Un certain nombre de types de documents sont hébergés dans le référentiel BOFiP. Le type de ceux-ci est précisé dans la rubrique *dataType* du paragraphe part du fichier « *document.xml* ». De la même manière, la rubrique *dataRef* précise le nom attribué à ce document source.

Suivant le type de document, on trouvera les documents sources suivants :

Type de document	Fichier source
Fichier	data.dat data.doc data.ods data.odt data.txt data.xls data.xml
Contenu	data.html
Image	data.gif data.jpeg data.jpg data.png
PlanClassement	data.xml

6 ÉVOLUTIONS 2020

Types de documents

Dans les noms de dossiers ainsi que dans les références de document .xml, les libellés des catégories ont évolué ainsi :

anciennement	maintenant
1-commentaire	Commentaire
2-bareme	Barème
3-formulaire	Formulaire
4-lettre-type modele	Lettre Type Modèle
5-cartographie	Cartographie
7-autres annexes	Autres annexes

Format des documents de type Contenu

Les données de type *contenu doctrinal* sont au format html et non plus xml.

Type des relations contenu – annexe

Les relations entre des documents de contenu doctrinal et leurs annexes sont de type « references » et non plus « requires » :

```
<dc:relation type="references">Contenu.Autres annexes:2407-PGP</dc:relation>
```