

**OpenSirene**

# 1. Nommage

Donner un nom explicite aux fichiers générés :

```
sirc-266_266_13705_201606_L_P_20161010_121909418.csv
```

VS. 

```
sirene_stock_2016-06-01_utf8.csv
```

**Gain : meilleure compréhension pour le réutilisateur.**

## 2. Encodage

Utiliser UTF-8 pour tous les fichiers avec [iconv](#) :

```
iconv -f latin1 -t utf-8 sirc.csv > sirc-utf8.csv
```

**Gain : prise en main rapide par le réutilisateur.**

### 3. Dates

Utiliser le format recommandé de l'ISO 8601 pour les dates :

20140108 vs 2014-01-08

**Gain : parsing facilité pour le réutilisateur.**

## 4. CSV

Utiliser la virgule (,) comme délimiteur et les guillemets (") que si nécessaire avec [csvformat](#) :

```
csvformat sirc.csv -D ',' -d ';' 
```

**Gain : fichier plus petit et parsing facilité.**

## 5. Éditorialisation

Rendrent explicites les choix réalisés lors de la génération des fichiers :

```
2016-10-01 : le contenu de la colonne "EMAIL" est supprimé  
pour éviter les courriers indésirables
```

**Gain : meilleure compréhension pour le réutilisateur.**

## 6. Normalisation

Ne pas utiliser des labels pour les catégories mais des identifiants courts et produire la documentation associée :

10 à 19 salariés vs. 11

**Gain : fichier plus petit.**

# 7. Redondance

Supprimer les informations (colonnes) en double dans les fichiers :

**TEFEN** vs. **LIBTEFEN** par exemple.

**Gain : fichier plus petit.**



## 8. Production

Expliquer les contraintes dans la génération de ces données :

- dans quel cadre sont-elles actuellement utilisées ?
- quels outils sont mis en œuvre dans la génération ?
- quel est l'interlocuteur pour améliorer la production ?

**Gain : meilleure compréhension et réduction des frustrations pour le réutilisateur.**

## 9. Erreurs

Rendre publique la liste des problèmes identifiés en amont et en cours de correction :

- simple fichier texte
- gestionnaire de ticket
- etc

**Gain : réduction des frustrations pour le réutilisateur.**

# 10. Changelog

Fournir un document succinct permettant d'être tenu informé des évolutions du format dans le temps :

2016-11-15 : retrait de la colonne redondante "LIBTEFEN"

**Gain : meilleure compréhension et réduction des frustrations pour le réutilisateur.**