

# La qualité de mes données

- Le jeu des 20 erreurs
- Les grands types de problèmes
  - Outils et ressources

# Le jeu des 20 erreurs

Prénom	Nom	Email	Tel mobile	Tel fixe	Code postal
Clément	MONTAIGNE	clmont@gmail.com	06.45.65.15.89		14200
SAND	Anouk	assand@free.fr	622405584	05 13 72 72 37	13012
JOEY	ABBIDI	abbidi.joey@gmail.com			69100
N\dao	AFFAR	yaffar@orange.com	06 96 76 48 87	02.14.65.47.96	75002
Charles	Népote	charles.nepote@fing.org	06 16 72 72 38	03 25 67 67 38	13012
Pascal	Berger	pascal.berger.laposte.net			74112
Jacques	Bergerac (de)	jacques-de-bergerac@free.fr	06 96 45 21 58		74300
Lucy	de Belt	debelt-lucy@wanadoo.fr		04 45 68 66 54	69600
Franck	LARRIVEE	franck.larrivee@cegetel.net			
Émilie	Desrêches	desrêches-emilie@sfr.com	+33 6 96 61 75 36		
VINCENT	Droz	vincent.droz@gmail.com			65140
Emile	Le Borgne	emile@leborgne.fr		04 76 60 46 23	7100
Hugues	Brigaud	hbrigaud@orange.com		01 40 73 20 01	
Juliette	Andrassi	andrassi@free.fr	06-29-48-86-87		
Paul	Borges	pb@justice.gouv.fr	07 87 69 54 64		75008
Melinda	Braque	melinda_braque@gmail.com			
Hervé	CHOMAT	herve13000@free.fr			10300
Philippine	Destrée	pdestree@caramail.fr	06 48 72 39 11	09 99 08 07 05	
Camille	Astruc	cam-truc@laposte.net	06 94 6è 28 32		17200
Françoise	Bano	frabano@wanadoo.fr		03 76 87 34 09	111120
Diego	de la Vega	zorro@zorro.xxx	06 53 67 43 12		
Dork	Bot	dorkbot+ventprivee@blackmail.com			27500
Basile	Bouas	basile@bouas@numericable.fr	06 73 85 58 90		

- Problèmes syntaxiques et liés à la forme (sigles, noms propres, codes...)
- Problèmes sémantiques (contenu/validité des contenus, données périmées...)
- Manque de données ou surabondance de données
- Problèmes liés à la réglementation, les conventions d'usage, l'éthique
- Problèmes liés à l'environnement des données : fichiers, encodage, métadonnées
  - accessibilité des données (licence, droit d'accès), caractéristiques du fichier (pb de format), documentation, pb liés aux normes, standards, codes..., design/production des données, gouvernance des données (absence de remontée d'erreur, contrôle qualité des données...)

<https://infolabs.io/qualite-donnees>

Des outils pour améliorer a posteriori

Les tableurs (tri, rechercher&remplacer, filtre)

Des logiciels, applications en ligne : [transformy.io](https://transformy.io), [openrefine](https://openrefine.org)...

En amont : documenter, « contraindre » les saisies, anonymiser, animer la démarche qualité

Des ressources de l'infolab de la Fing :

Guide méthodologique en ligne :

« [Qualité des données, petit guide collaboratif de survie. Fing, 2018](#) »

La page Infolab qualité : <https://infolabs.io/qualite-donnees>

Exemple de [rapport de qualité](#) et jeu de données étudié

Un outil d'animation, le sprint qualité <https://infolabs.io/sprint-qualite>