

A photograph of two women in a modern office setting. One woman, wearing a red sweater, is seated and looking at a computer monitor. The other woman, with her hair in braids, is standing behind her, leaning over the desk and smiling as they look at the screen together. The background shows a blurred office environment with desks and equipment.

EXPERIAN  
INNOVATION  
SUMMIT  
2019

# DATA ENGINEERING : NOUVEAU MOTEUR DE CROISSANCE

DATATECHNOLOGYANALYTICS





# Aperture Studio : une vision client à 360° pour de meilleurs résultats

David Zydron  
Experian



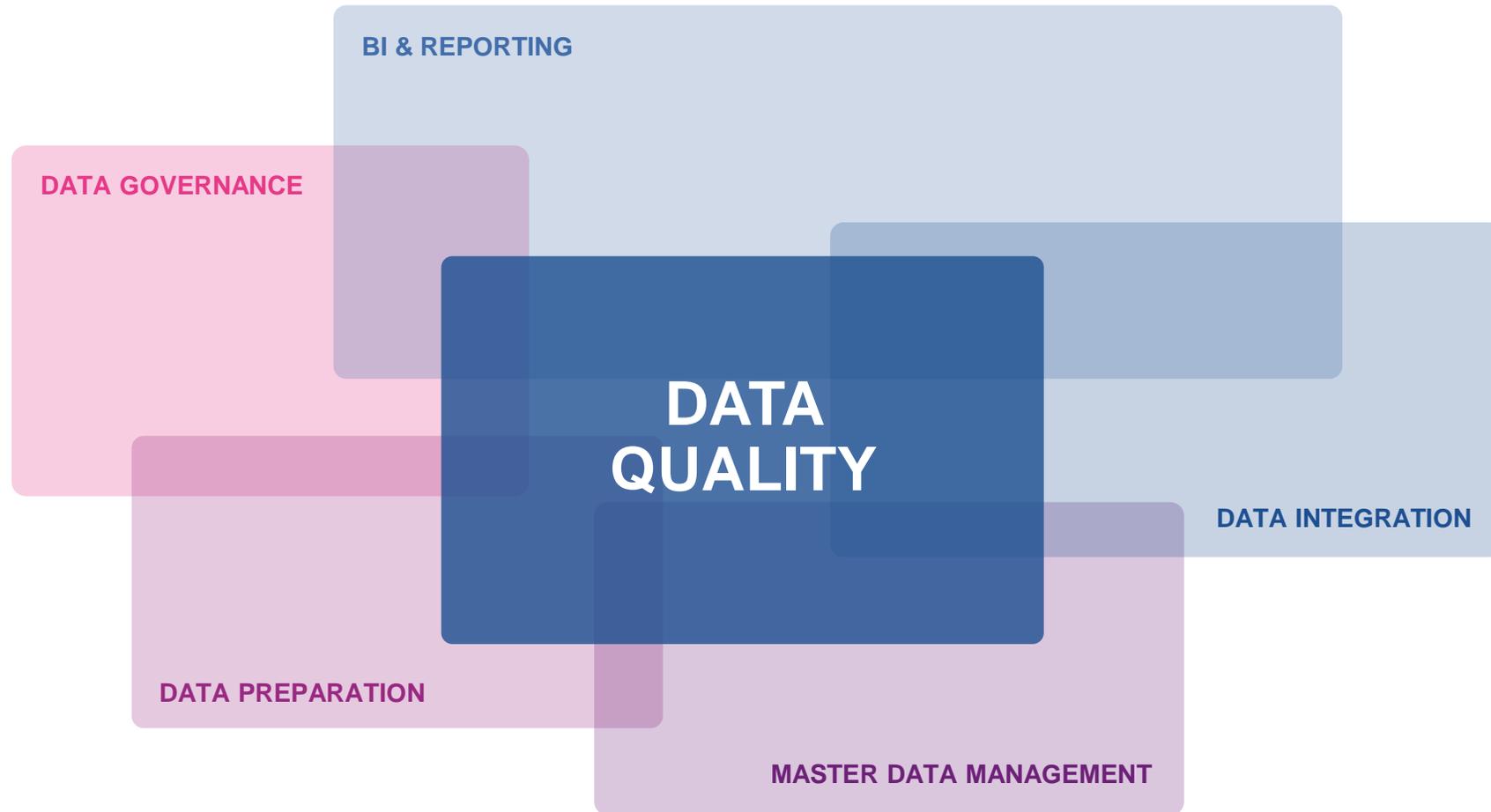


# Aperture Data Studio

Une vision client à 360° pour de meilleurs résultats



# L'importance de la donnée



# Problèmes identifiés chez nos clients



Législation



Vue unique du client



Architecture des données



Gouvernance des données



Qualité des données

# Méthodologie Experian



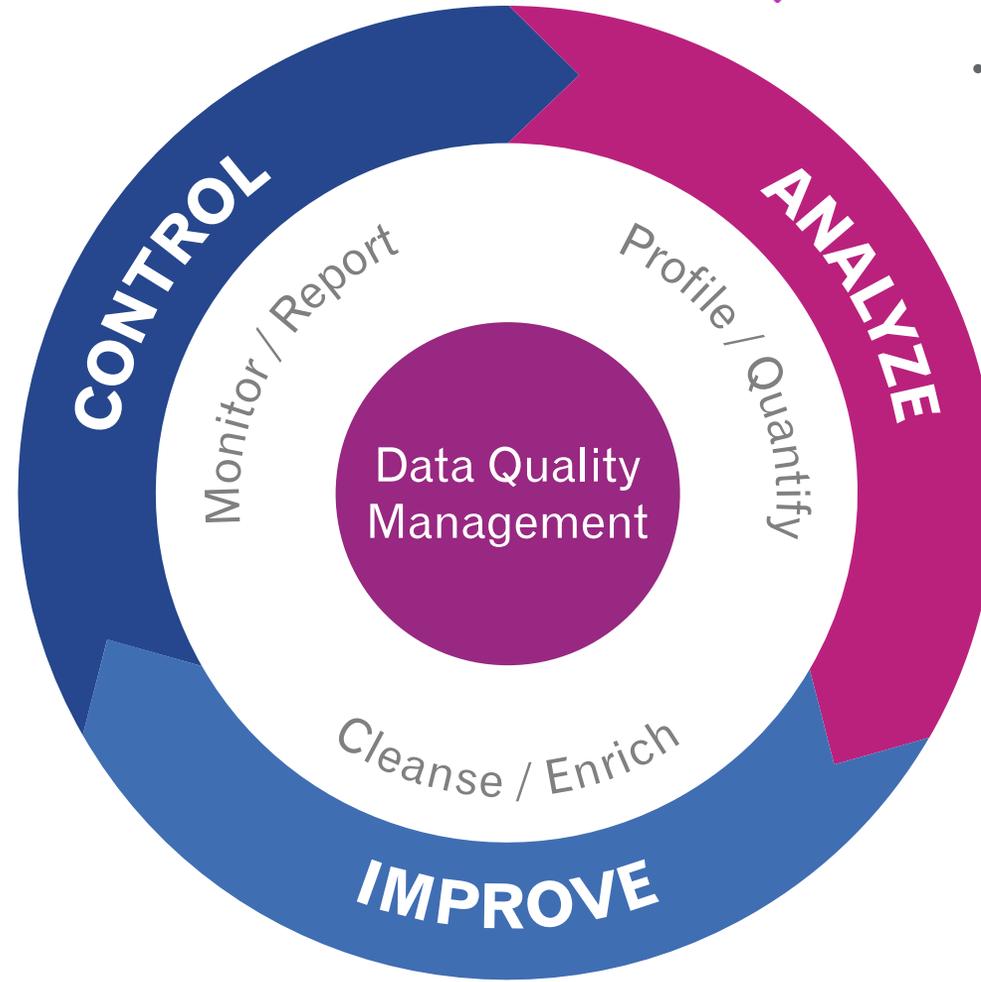
## Analyser

- **Profiler:** Rechercher, cataloguer, découvrir les inconnus des inconnus
- **Evaluer:** Mesurer la qualité de la donnée, analyser la cause principale
- **Quantifier:** Définir et prioriser un impact métier



## Contrôler

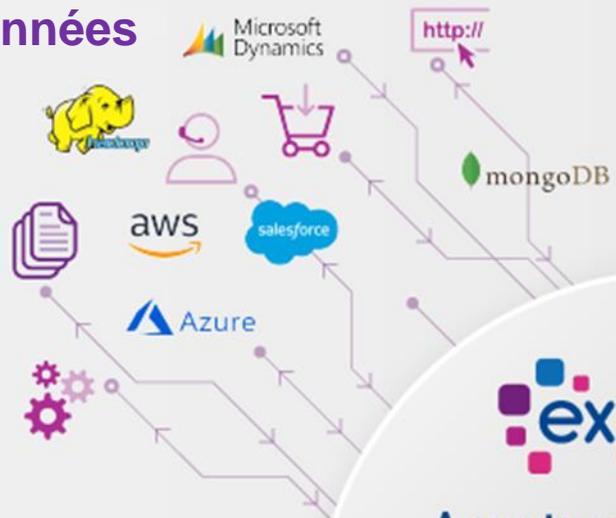
- **Déployer:** Implémenter des règles métier de qualité de la donnée
- **Signaler:** Mesurer les indicateurs métier
- **Suivre:** Mesurer la qualité de la donnée au cours du temps



## Améliorer

- **Transformer:** Nettoyer, fusionner et standardiser
- **Enrichir:** Intégrer des données de référence
- **Prototyper:** Créer et valider de façon dynamique les améliorations

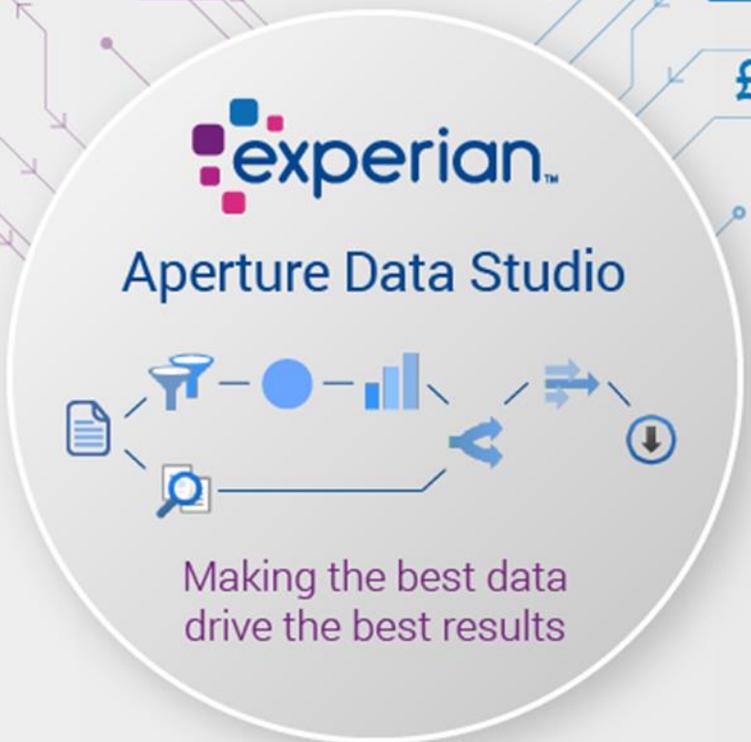
Vos Données



Donnée D'Experian



Connecter



Construire



Réponse à votre problématique





# Nous avons 3 systèmes différents contenant des données clients

- Système interne utilisé par un centre d'appels
- Données existantes de qualité mixte



CRM



- Données de compte en ligne saisies directement par les clients
- Données de transactions et informations de commandes



eCommerce



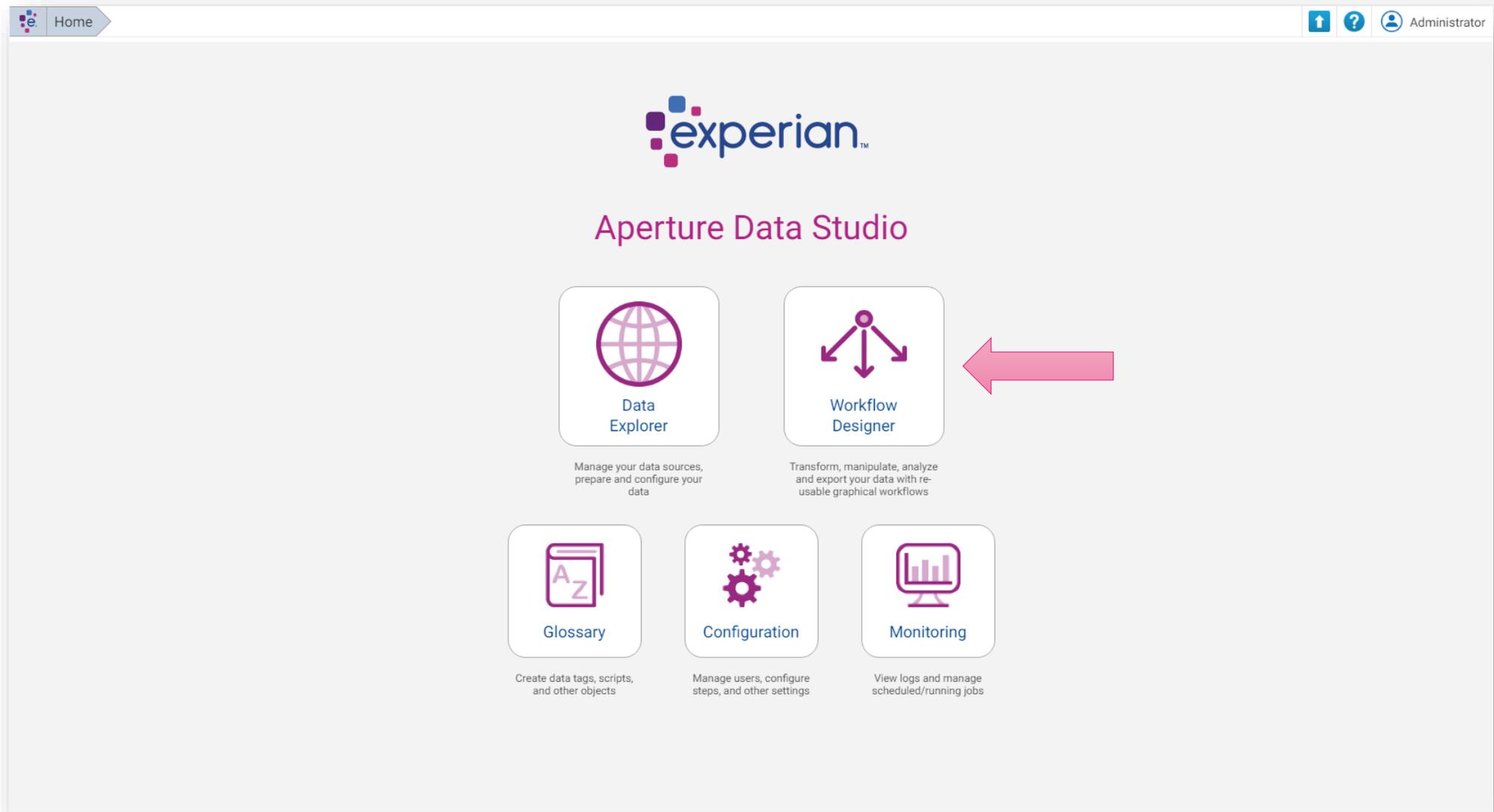
- Fichier Excel du programme de fidélité
- Données saisies par les clients sur des formulaires papier, puis saisies manuellement par le personnel



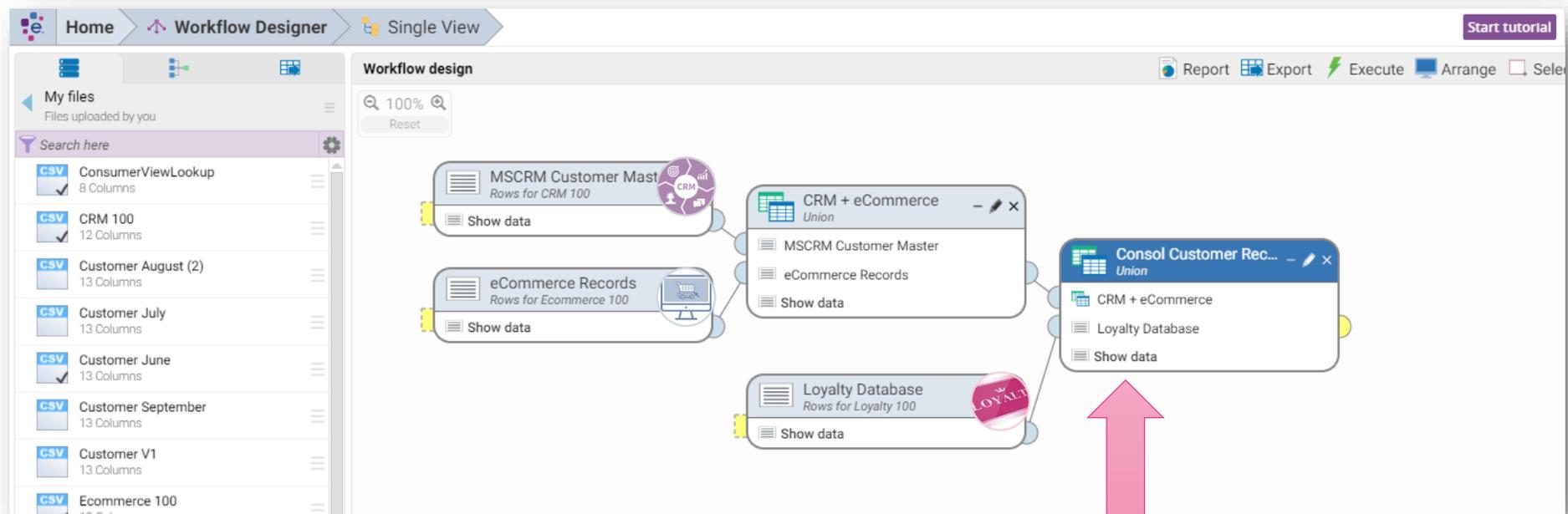
Loyalty Database



# Construisons ensemble !



# Concevoir un workflow via un système de glisser/déposer en utilisant des composants dans l'interface utilisateur web



En avançant, je peux toujours voir ce qu'il advient des données

# Voici mon jeu de données joint

Single View Rows for Consol Customer Rec (Con... Start tutorial ? Administrator

Rows for Consol Customer Rec (Consol Customer Rec (1)) Download as CSV Options X Close

|    | Record Id | Title | First Name | Last Name | Email                 | Gender | Street Address        | City         | State      | Zip Code | Country       | Phone         | Dob         | Last Updated Days | Source    | Points | Spend |
|----|-----------|-------|------------|-----------|-----------------------|--------|-----------------------|--------------|------------|----------|---------------|---------------|-------------|-------------------|-----------|--------|-------|
| 1  | C-08567   |       | Alberto    | Aasaf     | aaasaf9t@cdbaby.c...  | Male   | 857 Columbus Way      | Vancouver    | Washington | 98682    | United States | 1-(360)877... | 02-May-1967 | 230               | CRM       |        | 2435  |
| 2  | C-07309   | Mr    | Wally      | Abad      | wallya790@flavors.... |        | 839 Oakridge Junction | Saint Louis  | Missouri   |          | United States | 1-(314)447... | 01-Jan-2001 | 88                | CRM       |        | 2503  |
| 3  |           | Mr    | Walter     | Abad      |                       | Male   | 839 Oakridge Junction | Saint Louis  | Missouri   |          | United States | 1-(314)447... | 01-Jan-2001 | 169               | CRM       |        | 14    |
| 4  |           | Mr    | Walter     | Abad      | wallya790@flavors.... |        |                       | St Louis     | Missouri   |          | United States | 1-(314)447... | 01-Jan-2001 | 941               | Loyalty   | 901    | 1413  |
| 5  |           | Mr    | Walter     | Abad      | wallya790@flavors.... |        | Oakridge Junction     | Saint Louis  | Missouri   |          | United States | 1-(314)447... | 01-Jan-2001 | 855               | eCommerce |        | 573   |
| 6  | C-24223   |       | Lora       | Abbatini  | l.abbatini31@cdc.g... |        | 0 Katie Place         |              | Illinois   | 62711    | United States | 1-(217)746... | 31-Jan-1954 | 498               | CRM       |        | 389   |
| 7  | C-8207    | Ms    | Laura      | Abbatini  | l.abbatini31@cdc.g... |        | 10 Katie Place        | Springfield  |            | 62711    | United States | 1-(217)746... | 31-Jan-1954 | 805               | eCommerce |        | 1730  |
| 8  | C-24223   |       | Laura      | Abbatini  | l.abbatini31@cdc.g... |        | 0 Katie Place         |              |            | 62711    | United States | 1-(217)746... | 31-Jan-1954 | 535               | Loyalty   | 2227   | 858   |
| 9  | C-23899   |       | Phebe      | Abbay     |                       | Female | 9 Brickson Park Road  | Indianapolis | Indiana    | 46207    | United States | 1-(317)764... | 22-Jul-1970 | 854               | CRM       |        | 325   |
| 10 | C-23899   | Ms    | Phoebe     | Abbay     | pabbs@ricoh.com       |        |                       | Indianapolis | Indiana    | 46207    | United States | 1-(317)764... | 22-Jul-1970 | 849               | eCommerce |        | 1787  |
| 11 | C-0017    |       | Phoebe     | Abbay     | pabbs@ricoh.com       | Female | 9 Brickson Park Road  |              |            | 46207    | United States | 1-(317)764... | 22-Jul-1970 | 790               | Loyalty   | 100    | 938   |

# Voici mon jeu de données joint

Single View Rows for Consol Customer Rec (Con...)

Rows for Consol Customer Rec (Consol Customer Rec (1))

|    | Record Id | Title | First Name | Last Name | Email                 | Gender | Street Address        | City         | State      | Zip Code | Country       |
|----|-----------|-------|------------|-----------|-----------------------|--------|-----------------------|--------------|------------|----------|---------------|
| 1  | C-08567   |       | Alberto    | Aasaf     | aaasaf9t@cdbaby.c...  | Male   | 857 Columbus Way      | Vancouver    | Washington | 98682    | United States |
| 2  | C-07309   | Mr    | Wally      | Abad      | wallya790@flavors.... |        | 839 Oakridge Junction | Saint Louis  | Missouri   |          | United States |
| 3  |           | Mr    | Walter     | Abad      |                       | Male   | 839 Oakridge Junction | Saint Louis  | Missouri   |          | United States |
| 4  |           | Mr    | Walter     | Abad      | wallya790@flavors.... |        |                       | St Louis     | Missouri   |          | United States |
| 5  |           | Mr    | Walter     | Abad      | wallya790@flavors.... |        | Oakridge Junction     | Saint Louis  | Missouri   |          | United States |
| 6  | C-24223   |       | Lora       | Abbatini  | l.abbatini31@cdc.g... |        | 0 Katie Place         |              | Illinois   | 62711    | United States |
| 7  | C-8207    | Ms    | Laura      | Abbatini  | l.abbatini31@cdc.g... |        | 10 Katie Place        | Springfield  |            | 62711    | United States |
| 8  | C-23899   |       | Laura      | Abbatini  | l.abbatini31@cdc.g... |        | 0 Katie Place         |              |            | 62711    | United States |
| 9  | C-23899   |       | Phebe      | Abbey     |                       | Female | 9 Brickson Park Road  | Indianapolis | Indiana    | 46207    | United States |
| 10 | C-23899   | Ms    | Phoebe     | Abbey     | pabbs@ricoh.com       |        |                       | Indianapolis | Indiana    | 46207    | United States |
| 11 | C-0017    |       | Phoebe     | Abbey     | pabbs@ricoh.com       | Female | 9 Brickson Park Road  |              |            | 46207    | United States |

Annotations:

- Données manquantes (points to empty Title field in row 8)
- Noms courts et fautes d'orthographe (circles "Lora" and "Phoebe" in rows 6 and 10)
- Adresses incomplètes et incorrectes (circles "0 Katie Place" and "9 Brickson Park Road" in rows 6, 8, 9, 10)

Cet extrait a un certain nombre de problèmes que nous voyons généralement chez nos clients...

Ces problèmes ne concernent pas uniquement la  
déduplication

C'est aussi une question d'harmonisation

Il faut préserver les enregistrements utiles

Et décider intelligemment de supprimer les informations  
inutiles

# Mais d'abord... comment puis-je savoir que les données reflètent la réalité?

The screenshot shows the 'Workflow Designer' interface in 'Single View' mode. The workflow consists of three steps connected in a sequence:

- Step 1:** 'Rows for Consol Customer Rec (Consol)' with a 'Show data' button.
- Step 2:** 'Validate addresses' (Powered by Experian Batch). It shows 'Selected columns' (4 / 17) and 'United States of America' as a filter. It has 'Show data' and 'Additional options' buttons.
- Step 3:** 'Validate emails'. It shows a 'Show data' button and a 'Domain level' filter.

The interface includes a 'Workflow steps' sidebar on the left with options like 'Analyze trends', 'Branch', 'Chart', 'Experian Link Demo', 'Export', and 'Filter'. The top navigation bar includes 'Home', 'Workflow Designer', 'Single View', and a 'Start tutorial' button. The main toolbar contains 'Report', 'Export', 'Execute', 'Arrange', and 'Select'.

Par exemple Aperture vous permettra de vérifier les enregistrements postaux et les domaines de messagerie.



# Maintenant que nous avons des données cohérentes, trouvons les doublons

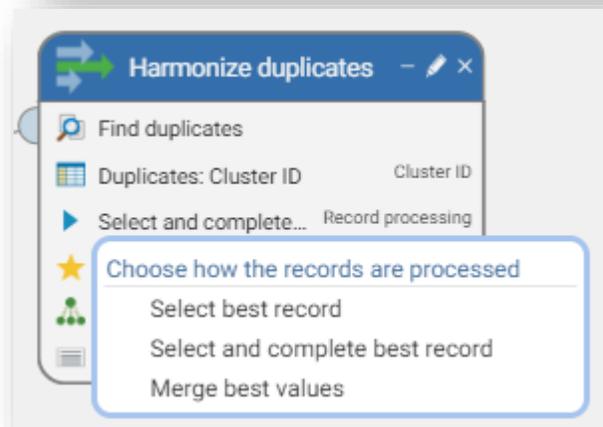
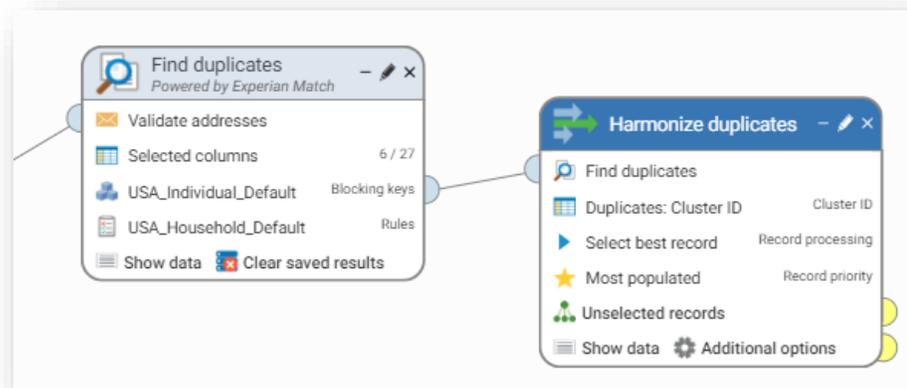
The screenshot shows two panels from a data tool. The left panel, titled 'Validate emails', has a green checkmark icon and options for 'Email', 'Domain level', and 'Show data'. The right panel, titled 'Find duplicates' and 'Powered by Experian Match', has a magnifying glass icon and options for 'Validate emails', 'Selected columns' (8 / 29), 'USA\_Individual\_Default' (Blocking keys), 'USA\_Individual\_Default' (Rules), 'Show data', and 'Clear saved results'.

|    | Title | First Name | Last Name | Email         | Dupl... | Gender | Street Address           | City      | State  | Zip Code | Phone      | Last Updated Days | Source    | Points |
|----|-------|------------|-----------|---------------|---------|--------|--------------------------|-----------|--------|----------|------------|-------------------|-----------|--------|
| 39 |       | Fernanda   | Biddulph  | fbiddulphc... | 13      |        | 1654/1 Forest Run Avenue | Las Ve... | Nevada | 89120    | 7708043363 | 760               | CRM       |        |
| 40 | Miss  |            | Biddulph  | fbiddulphc... | 13      |        | 1654 Forest Run Avenue   | Las Ve... | NA     | 89120    | 7708043363 | 555               | eCommerce |        |
| 41 |       | Fernanda   | Biddulph  | fbiddulphc... | 13      | Female | 1654 Forest Run Ave      | Las Ve... |        | 89120    | 7708043363 | 129               | Loyalty   | 33     |

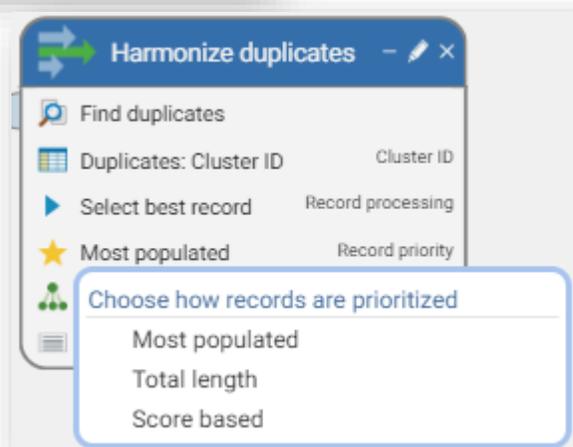
Bien que ces enregistrements ne soient pas identiques, le fuzzy matching leur a donné le même identifiant de cluster.

Je peux utiliser ces données pour sélectionner les **Golden records**

# Cette étape est appelée “Harmonisation”

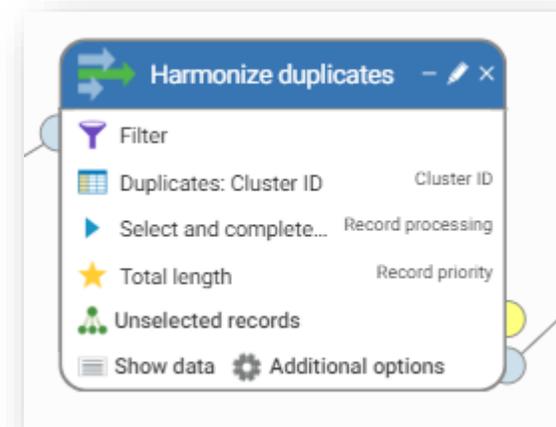


Tout d'abord, je choisis les enregistrement que je veux conserver.



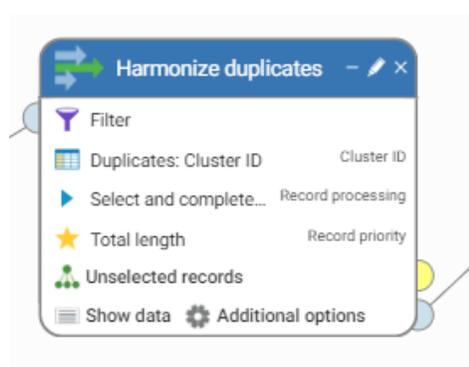
Ensuite, je décide ce que “Best Record” signifie

(NB - Je pourrais utiliser la méthode «Score» ici sur le champ «Dernière mise à jour» pour sélectionner uniquement les enregistrements les plus récents.)



Vue finale du processus de dédoublonnage

# Le Résultat



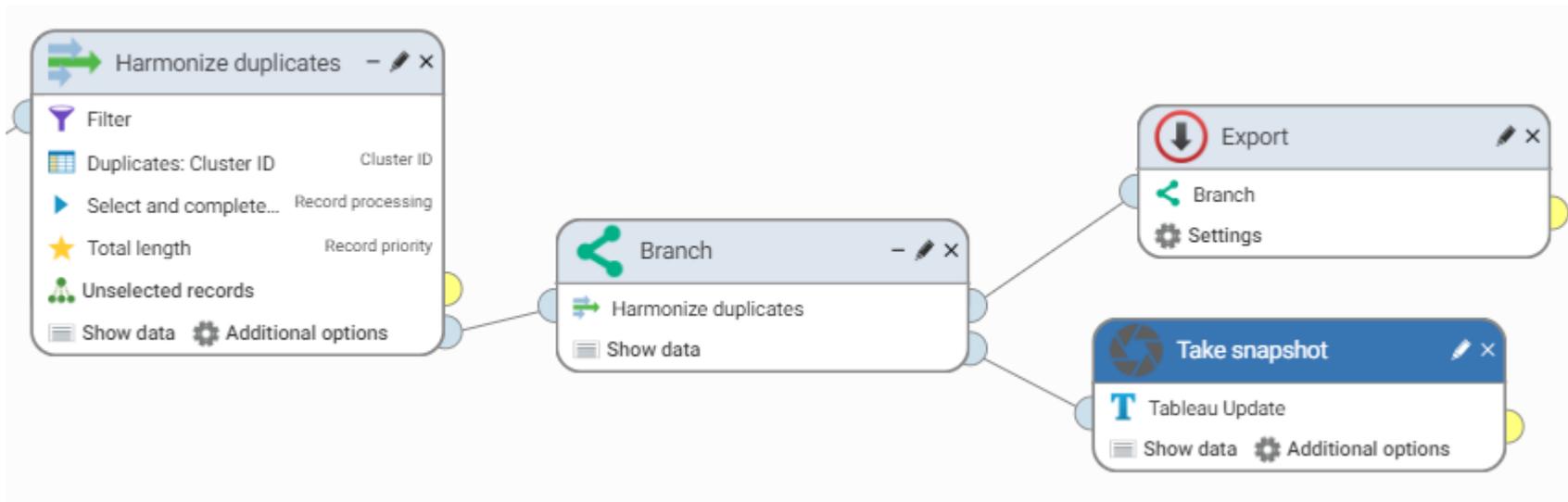
On part de...

|    | Title | First Name | Last Name | Email         | Dupl... | Gender | Street Address           | City      | State  | Zip Code | Phone      | Last Updated Days | Source    | Points |
|----|-------|------------|-----------|---------------|---------|--------|--------------------------|-----------|--------|----------|------------|-------------------|-----------|--------|
| 39 |       | Fernanda   | Biddulph  | fbiddulphc... | 13      |        | 1654/1 Forest Run Avenue | Las Ve... | Nevada | 89120    | 7708043363 | 760               | CRM       |        |
| 40 | Miss  |            | Biddulph  | fbiddulphc... | 13      |        | 1654 Forest Run Avenue   | Las Ve... | NA     | 89120    | 7708043363 | 555               | eCommerce |        |
| 41 |       | Fernanda   | Biddulph  | fbiddulphc... | 13      | Female | 1654 Forest Run Ave      | Las Ve... |        | 89120    | 7708043363 | 129               | Loyalty   | 33     |

A...

|   | Title | First Name | Last Name | Email         | Du... | Gender | Street Address           | City      | State  | Zip Code | Phone      | Last Updated Days | Source | Points |
|---|-------|------------|-----------|---------------|-------|--------|--------------------------|-----------|--------|----------|------------|-------------------|--------|--------|
| 1 | Miss  | Fernanda   | Biddulph  | fbiddulphc... | 13    | Female | 1654/1 Forest Run Avenue | Las Ve... | Nevada | 89120    | 7708043363 | 760               | CRM    | 33     |

# Et enfin... ajouter une ou deux étapes finales au workflow



- Mettez à jour vos enregistrements source avec les données épurées
- Mettez à jour votre MDM
- Mettez à jour votre système de BI

A photograph of two women in a modern office setting. One woman, with dark hair in braids, is leaning over a computer monitor and pointing at the screen. The other woman, wearing a red sweater, is looking at the monitor with a focused expression. The background shows office desks and equipment.

EXPERIAN  
INNOVATION  
SUMMIT  
2019

**DATA ENGINEERING :**  
**NOUVEAU MOTEUR**  
**DE CROISSANCE**

DATATECHNOLOGYANALYTICS

