

Référentiels de données

Note de cadrage - mars 2026

Le Groupe CCI Paris Ile-de-France

La CCI Paris Ile-de-France est un établissement public de l'Etat, composé de 8 CCI départementales et territoriales et d'Etablissements d'enseignement supérieur consulaires :



La CCI Paris Ile-de-France est présente sur l'ensemble du territoire francilien. Elle met ses services et ses compétences à disposition des entreprises, des jeunes et des salariés, et des collectivités.

Cet engagement s'articule autour de 4 missions fondamentales :

- Accompagner les entreprises et commerces dans leur développement et leur quotidien
- Former les jeunes et les salariés aux métiers d'aujourd'hui et de demain
- Favoriser l'attractivité du territoire au travers de son activité congrès & salons
- Représenter les entreprises auprès des acteurs publics pour promouvoir la mise en œuvre de mesures favorisant leur développement

Les enjeux et objectifs

Avec plus de 1,2 million d'entreprises sur son territoire, le système d'information de la CCI est de plus en plus riche et complexe. Il est composé d'un volume croissant de données issues de sources multiples. Ces données sont hétérogènes, non consolidées, souvent redondantes...

Dans ce contexte, le Data Office travaille à la définition d'une gouvernance des données et a engagé une réflexion sur la fiabilisation et le partage de la donnée au sein du Groupe.

Afin d'optimiser la gestion des données et mieux connaître ses clients, la CCI souhaite se doter d'un outil permettant de :

- Disposer d'une vue consolidée et centralisée des données de référence, constituant un point unique de vérité
- Partager et enrichir la donnée au sein de l'organisation
- Gérer et améliorer la qualité des données
- Mettre à disposition la donnée de manière sécurisée pour les équipes marketing, communication, finance, commerciales...

Les principaux référentiels

- Les entreprises d'Ile-de-France
- Les contacts et produits de l'appui aux entreprises
- Les contacts et formations du secteur Education
- Les adresses des tiers moraux et physiques
- Les collaborateurs du Groupe

Référentiel	Nombre d'enregistrements (fourchette haute)
Contacts Appui	754 722
Produits Appui	2 076
Contacts Education	102 738
Formations Education	2 066
Entreprises	2 500 000
Etablissements	2 800 000
Dirigeants	3 000 000
Adresses	3 000 000
Collaborateurs	3 735
Total général	6 059 109

Les clients de la donnée

L'ensemble des directions et établissements sont susceptibles de générer et utiliser des données du SI, qu'il s'agisse de données spécifiques à leur activité ou de données transverses.

Les principaux clients sont les suivants :

- Le pôle Entreprises
- L'appui aux entreprises
- L'éducation
- Les EESC de la Chambre (Etablissements d'enseignement supérieur consulaires)
- La Direction International
- Le pôle Data CCIR Groupe
- Autres CCI (Essonne, Seine-et-Marne, Normandie...)
- La finance et la comptabilité

Les sources de données

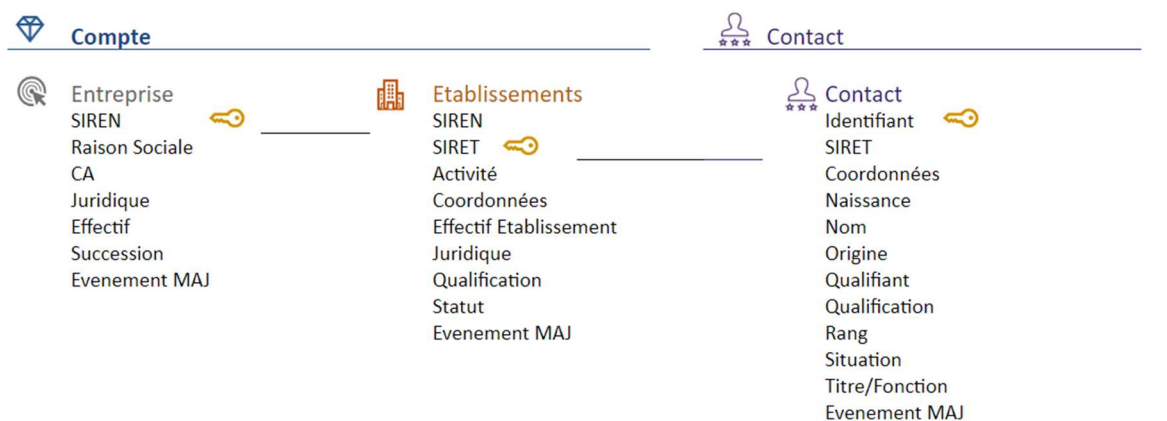
Les données peuvent provenir de sources internes ou externes, gratuites ou payantes, dont voici quelques exemples :

- Ressources internes :
 - Fichier Entreprises (pour les élections consulaires)
 - SIGR (Système d'Information Géographique Régional)
 - SAP (Données financières)
 - CRM (Marketing & Communication)
 - Yparéo (ERP Education)
 - HR 4YOU (Données RH)
- Ressources externes :
 - RNE (Registre National des Entreprises géré par l'INPI et opéré par CCI France)
 - INSEE, URSSAF, Datainfogreffe (Données entreprises)
 - Infolégale (Données légales et financières des entreprises et dirigeants)
 - DQE Software (Données clients)
 - Datagouv, douane.gouv.fr (Base Adresse Nationale, Données commerce extérieur...)

L'exemple du Référentiel Entreprises

Le modèle de données métier :

- 1 compte représente 1 entreprise et ses établissements,
- 1 entreprise est composée de 1 à n établissements,
- 1 établissement détient 1 à n contacts, dont 1 représentant légal principal
- 1 contact est unique dans la base, mais peut être associé à plusieurs établissements



Le cycle de vie de la donnée :

La donnée Entreprise évolue en permanence, et notamment au fil notamment des créations, modifications et radiations. Aujourd'hui, ces formalités sont réalisées sur le guichet unique, par le déclarant ou le mandataire de l'entreprise : <https://procedures.inpi.fr/>

Il est nécessaire de pouvoir modéliser/paramétrer le cycle de vie d'un objet ou et d'un attribut, à travers différents états, par exemple : créé, à vérifier, validé/qualifié, activé, à requalifier, obsolète, archivé...

Dans ce contexte, des workflows de validation doivent être définis et mis en œuvre :

- Workflows déclenchés selon les règles de gestion établies, paramétrables, pour une exécution en parallèle et/ou ordonnancée par étape
- Attribution des tâches en fonction de différents critères (rôle utilisateur, niveau d'habilitation, domaine d'intervention/ fonction, entité à laquelle l'utilisateur appartient...)
- Délégation de tâche par le validateur ou le superviseur
- Etc...

Besoins fonctionnels

Conception et évolution du modèle de donnée

- Créer de nouvelles entités/objets
- Ajouter/Supprimer/Modifier des attributs et types d'attribut (date, nombre, texte, liste de valeurs, lien, document, image, vidéo...)
- Modéliser des relations entre objets (clé primaire ou étrangère, association (bridge), hiérarchie de données (lien parents/enfants), héritage père/fils sur certains attributs

Acquisition de la donnée

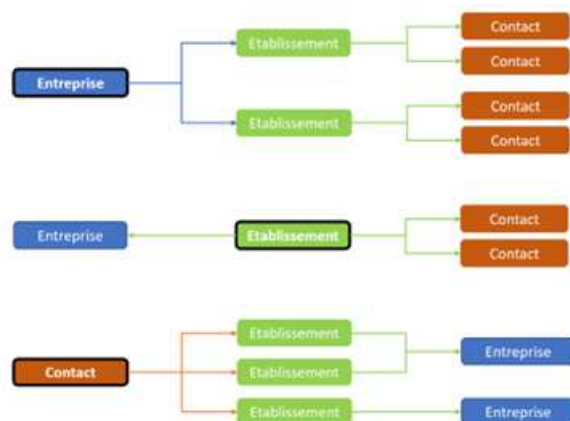
- Définir les workflows, paramétrer les règles, pour intégrer les données via API ou fichiers plats, en mode full/delta
- Piloter la gestion des anomalies, gérer les erreurs d'intégration

Gestion de la qualité de la donnée

- Garantir l'unicité des objets métier (entreprise, produit, formation...) dans le référentiel
- Pouvoir définir quelle source de données est la plus fiable pour chaque attribut afin de piloter sa mise à jour (ex : données géographiques = source SIGR)
- Gérer la qualité de données (dédoublonnage, paramétrage de contrôles, mise en place de workflows pour la mise à jour des données, implémentation de règles de gestion...)

Accès à la donnée

- Accéder aux données afin de permettre des corrections ou reprises de données
- Accéder aux données via une IHM, en lecture ou écriture selon habilitations
- Créer des filtres pour rechercher les données, créer et partager des vues
- Extraire ou exposer la donnée via API
- Disposer de vues hiérarchiques :



Exposition de la donnée

- Exposer la donnée en mode base et en mode API, avec une gestion sécurisée des données
- Mettre à disposition la donnée en temps réel/consommation dynamique pour l'usage des ERP ou d'autres outils

Administration et sécurité

- Disposer d'une console de paramétrage et de pilotage de la solution (administrer les rôles et droits, paramétrer les workflows, archiver/historiser des données (tables et/ou attributs particuliers...))
- Sécurité appliquée au niveau de la ligne et/ou de l'attribut
- Garantir le respect du RGPD :
 - Définir la durée de conservation de la donnée (jusqu'au niveau de l'attribut),
 - Définir les règles de gestion pour anonymiser, supprimer, pseudonymiser certaines données au terme de leur durée de vie
 - Restreindre l'accès à certaines données pour des rôles/donner des habilitations spécifiques pour visualiser certains attributs
 - Pouvoir paramétrer l'outil pour garantir la conformité par rapport au RGPD