

**PROGRAMME DE FORMATION DÉVELOPPEUR EN INTELLIGENCE ARTIFICIELLE  
OU CERTIFICATION PROFESSIONNELLE DÉVELOPPEUR EN INTELLIGENCE  
ARTIFICIELLE ET DATA SCIENCE RNCP36581**

Bloc E6.1 – Créer un modèle de données d'une solution I.A en utilisant des méthodes  
de Data science

Cahier des Charges de la MSPR « Mise en œuvre d'un processus ETL répondant au  
besoin d'une solution I.A »

**COMPÉTENCES ÉVALUÉES :*****Certification professionnelle Développeur en intelligence artificielle et data science RNCP36581 :***

- Définir les sources et les outils nécessaires pour permettre de collecter les données.
- Recueillir de manière sécurisée les informations à partir de sources adaptées (sources hétérogènes, internes fournies par le client ou externes accessibles en Open Data) permettant de définir les données à collecter pour réaliser l'architecture de données.
- Paramétrer les outils afin d'importer les données de manière automatisée et sécurisée.
- Analyser, nettoyer, trier et s'assurer de la qualité des données afin de les rendre exploitables pour la solution I.A, en utilisant des outils d'analyse et de visualisation des données et se basant sur des approches de la Data science.
- Construire la structure de stockage des données (modèle de données) qui répond au mieux au besoin d'analyse.
- Représenter graphiquement les relations entre les données afin de les visualiser en créant des tableaux de bord accessibles à tout public garantissant ainsi l'accessibilité numérique.
- Exploiter de manière automatisée et analyser les informations recueillies dans les structures de stockage des données (requêtage ou interrogation) afin de répondre aux exigences de la solution IA défini dans le cahier des charges.

***Programme de formation Développeur en intelligence artificielle :***

- Automatiser l'extraction de données depuis un service web, une page web (scraping), un fichier de données, une base de données et un système big data en programmant le script adapté afin de pérenniser la collecte des données nécessaires au projet.
- Développer des requêtes de type SQL d'extraction des données depuis un système de gestion de base de données et un système big data en appliquant le langage de requête propre au système afin de préparer la collecte des données nécessaires au projet.
- Développer des règles d'agrégation de données issues de différentes sources en programmant, sous forme de script, la suppression des entrées corrompues et en programmant l'homogénéisation des formats des données afin de préparer le stockage du jeu de données final.
- Créer une base de données dans le respect du RGPD en élaborant les modèles conceptuels et physiques des données à partir des données préparées et en programmant leur import afin de stocker le jeu de données du projet.
- Développer une API mettant à disposition le jeu de données en utilisant l'architecture REST afin de permettre l'exploitation du jeu de données par les autres composants du projet.

## PHASE 1 : PRÉPARATION DE CETTE MISE EN SITUATION PROFESSIONNELLE RECONSTITUÉE

- **Durée de préparation** : 19 heures
- **Mise en œuvre** : Travail d'équipe constituée de 4 apprenants (5 maximum si groupe impair)

## PHASE 2 : PRÉSENTATION ORALE COLLECTIVE + ENTRETIEN COLLECTIF

- **Objectif** : mettre en avant et démontrer que les compétences visées par ce bloc sont bien acquises.
- **Moyen** : L'équipe utilise un support de présentation
- **Durée totale par groupe** : 50 mn se décomposant comme suit :
  - 20 mn de soutenance orale par l'équipe.
  - 30 mn d'entretien collectif avec le jury (questionnement complémentaire).
- **Jury d'évaluation** : 2 personnes (binôme d'évaluateurs) par jury – Ces évaluateurs ne sont pas intervenus durant la période de formation et ne connaissent pas les apprenants à évaluer.

## I – CONTEXTE

**ObRail Europe** est un observatoire indépendant spécialisé dans le ferroviaire et la mobilité durable. Créée en 2018, l'organisation s'est imposée comme un acteur de référence dans l'analyse des flux ferroviaires européens et la promotion du transport bas-carbone.

Sa mission principale est de :

- Collecter et analyser les données relatives aux dessertes ferroviaires en Europe.
- Évaluer l'impact environnemental des modes de transport longue distance.
- Produire des études comparatives destinées aux décideurs politiques et aux opérateurs ferroviaires.
- Accompagner la transition écologique en favorisant l'usage du train comme alternative à l'avion sur les trajets intra-européens.

ObRail travaille en partenariat avec :

- Les institutions européennes (Commission européenne, Parlement européen) pour alimenter les politiques publiques.
- Des ONG environnementales telles que Transport & Environnement ou Back-on-Track afin de renforcer l'argumentaire écologique.
- Les opérateurs ferroviaires (SNCF, ÖBB Nightjet, DB, Trenitalia) qui souhaitent valoriser leurs services de trains de jour et de nuit.

L'organisation s'inscrit dans les grandes stratégies européennes :

- Le Green Deal européen, qui vise la neutralité carbone à l'horizon 2050.
- Le programme TEN-T (Trans-European Transport Network), qui a pour objectif de renforcer l'intégration et la compétitivité du réseau ferroviaire en Europe.

**ObRail** souhaite comparer la contribution des trains de jour et des trains de nuit au maillage ferroviaire européen. L'objectif est de **mesurer leur rôle dans la construction d'une mobilité durable** et d'évaluer leur potentiel en tant qu'alternative crédible à l'avion sur les trajets intra-européens.

Pour cela, il est nécessaire de constituer **un référentiel de données fiable et harmonisé** qui servira :

- à l'analyse de la couverture ferroviaire,
- à l'entraînement de modèles d'IA,
- et au développement d'un service applicatif destiné aux partenaires.

L'étude préliminaire réalisée par ObRail Europe a mis en évidence plusieurs contraintes majeures auxquelles le projet devra répondre :

### 1. Dispersion des données

Les informations relatives aux dessertes ferroviaires européennes sont publiées par une multitude d'acteurs (opérateurs nationaux, plateformes open data, organismes statistiques). Chaque source utilise son propre format (CSV, flux GTFS, API, fichiers Excel, HTML scraping), ce qui rend leur intégration particulièrement complexe.

### 2. Qualité hétérogène des jeux de données

Les données collectées présentent des problèmes de complétude et de fiabilité : doublons, informations manquantes (codes de gares, fuseaux horaires, identifiants de lignes), incohérences dans les unités ou les formats de date/heure. Ces particularités doivent être traitées en amont afin de garantir la robustesse des analyses futures.

### 3. Absence de standardisation transfrontalière

Il n'existe pas de référentiel commun permettant de comparer facilement les dessertes ferroviaires entre pays européens. Chaque opérateur applique ses propres conventions (nomenclature des gares, structure des horaires, indicateurs de performance), ce qui empêche une comparaison homogène des données.

#### 4. Exigences réglementaires

Le projet doit se conformer au Règlement Général sur la Protection des Données (RGPD). Même si aucune donnée personnelle n'est traitée, les règles de sécurité, de documentation et de transparence dans la gestion des jeux de données doivent être respectées.

#### 5. Contraintes temporelles

ObRail doit livrer une première version exploitable de l'entrepôt de données aux institutions européennes avant la fin de l'année. Cette échéance impose un processus ETL fiable, automatisé et reproductible, permettant de garantir la mise à jour régulière des données

## II – CAHIER DES CHARGES

ObRail Europe souhaite disposer d'un **entrepôt de données unifié** permettant de mieux comprendre et comparer la contribution des trains de jour et des trains de nuit au maillage ferroviaire européen. Ce projet s'inscrit dans la stratégie de l'organisation d'appuyer les politiques publiques européennes en matière de mobilité durable et de fournir aux opérateurs ferroviaires comme aux ONG environnementales des données fiables et exploitables.

L'objectif principal est de mettre en place **un processus ETL automatisé et pérenne** permettant de collecter, transformer, stocker et exposer des données fiables sur les dessertes ferroviaires européennes. Cet entrepôt de données doit devenir **une source de référence** pour les analyses futures de ObRail, mais aussi pour ses partenaires institutionnels et associatifs.

Afin d'atteindre cet objectif général, plusieurs finalités opérationnelles doivent être poursuivies. Il s'agit notamment de :

- **Centraliser l'information** en agréant des données issues de multiples opérateurs ferroviaires et sources publiques.
- **Améliorer la qualité des données** en garantissant leur cohérence, leur homogénéité et leur traçabilité.
- **Faciliter l'analyse comparative** entre trains de jour et trains de nuit afin d'évaluer leur rôle dans la réduction des émissions de CO<sub>2</sub> et dans l'intégration du réseau ferroviaire européen.
- **Rendre les données accessibles** via une API REST documentée, pour en permettre l'utilisation par un public diversifié (data scientists, décideurs, ONG, grand public spécialisé).
- **Préparer les étapes suivantes** du projet global : entraînement de modèles prédictifs et mise en production d'une solution applicative.

La réussite du projet dépend également de l'implication et de la coordination entre les différentes parties prenantes. Les principaux acteurs identifiés sont :

- **Le commanditaire** : ObRail Europe, représenté par son département "Études et Innovation".
- **Les utilisateurs internes** : l'équipe Data Science de ObRail, chargée de développer les modèles IA.
- **Les utilisateurs externes** :
  - Institutions européennes, destinataires des analyses pour l'élaboration des politiques de mobilité.
  - ONG environnementales, utilisatrices des données pour sensibiliser le public et appuyer leurs plaidoyers.
  - Opérateurs ferroviaires, intéressés par une meilleure valorisation de leurs services.

Ce projet ne vise pas seulement à produire une base technique : il doit permettre à ObRail Europe de **renforcer sa crédibilité et son rôle d'acteur de référence** dans le débat sur la mobilité durable. Les livrables attendus doivent donc combiner **rigueur scientifique** (qualité et fiabilité des données) et **ouverture opérationnelle** (accessibilité, documentation, interopérabilité).

En fournissant un outil robuste et évolutif, ObRail pourra :

- alimenter ses études et publications,
- répondre rapidement aux sollicitations des institutions,
- et démontrer, chiffres à l'appui, que le rail – de jour comme de nuit – constitue une alternative compétitive et écologique au transport aérien.

### III – LES BESOINS EXPRIMÉS OBRail EUROPE

Afin d'atteindre les objectifs définis dans le cahier des charges, plusieurs contraintes fonctionnelles, techniques et réglementaires doivent être respectées. Ces éléments constituent les fondations du dispositif attendu.

#### Contraintes fonctionnelles

Le processus ETL devra permettre de :

- **Automatiser la collecte des données** issues de sources hétérogènes
- **Garantir la fiabilité et la traçabilité des informations collectées**, afin que les analyses produites puissent être exploitées par des tiers sans remise en question de leur qualité.
- **Nettoyer et transformer les données** pour les rendre exploitables : suppression des doublons, gestion des valeurs manquantes, homogénéisation des formats (codes gares, fuseaux horaires, unités de mesure).
- **Structurer les données** dans une base relationnelle adaptée aux besoins d'analyse, avec un modèle conceptuel et physique documenté.
- **Mettre à disposition les données via une API REST**, offrant des endpoints clairs, sécurisés et facilement exploitables par les utilisateurs finaux.
- **Proposer un tableau de bord de contrôle** permettant de suivre simplement la qualité des données et d'en vérifier la cohérence. Ce tableau de bord devra être accessible à tout public.

#### Contraintes techniques

Sur le plan technique, le dispositif devra respecter les exigences suivantes :

- **Interopérabilité** : le choix des outils et des langages (Python, R, SQL, frameworks pour l'API) est laissé libre, à condition que les livrables soient réutilisables et compatibles avec des environnements standards (Docker, Postman, PostgreSQL ou équivalent).
- **Automatisation et reproductibilité** : les scripts ETL doivent pouvoir être rejoués régulièrement afin d'intégrer de nouvelles données, sans nécessiter d'intervention manuelle lourde.
- **Qualité du modèle de données** : le schéma de la base devra permettre une exploitation efficace (requêtes, statistiques, analyses). Les relations entre les entités devront être claires et justifiées.
- **Accessibilité numérique** : la restitution des données via l'API devra respecter les standards d'accessibilité et permettre une prise en main rapide par des utilisateurs variés (data scientists, ONG, institutions).

#### Contraintes réglementaires et organisationnelles

Le projet devra être mené dans le respect du **Règlement Général sur la Protection des Données (RGPD)**. Même si les jeux de données ne contiennent pas d'informations personnelles, l'organisation exige que les principes de transparence, de documentation et de sécurisation soient appliqués.

Enfin, une **contrainte temporelle forte s'impose** : ObRail Europe doit livrer une première version exploitable de l'entrepôt de données dans un délai court, afin d'alimenter les travaux en cours du Parlement européen sur la mobilité durable. Cette échéance impose une démarche structurée, des livrables exploitables dès la première itération et une documentation claire.

## IV – LES LIVRABLES

À l'issue de la mission, ObRail Europe attend de l'équipe projet la remise d'un ensemble de livrables structurés, documentés et exploitables. Ces livrables constitueront à la fois la preuve du travail réalisé et la base opérationnelle pour les phases ultérieures du projet (exploitation par les data scientists, développement de modèles d'IA et mise en production d'applications).

### 1. Scripts ETL opérationnels

L'équipe devra produire un ensemble de scripts permettant l'automatisation du processus de collecte, de transformation et de chargement des données. Ces scripts devront être documentés, paramétrables et reproductibles, afin de pouvoir être rejoués régulièrement pour mettre à jour l'entrepôt de données.

### 2. Modèle conceptuel et physique des données

Un modèle conceptuel de données (MCD) et un modèle physique de données (MPD) devront être élaborés et fournis. Ils devront être accompagnés d'une justification des choix effectués (entités retenues, relations définies, contraintes d'intégrité). Ces modèles serviront de référence pour les évolutions futures de la base de données.

### 3. Base de données relationnelle alimentée

L'équipe projet devra livrer une base de données relationnelle opérationnelle (PostgreSQL ou équivalent), préremplie avec un jeu de données harmonisé issu du processus ETL. Cette base devra être exploitable immédiatement par les équipes de ObRail Europe.

### 4. API REST fonctionnelle

Une API REST devra être développée afin de mettre à disposition les données collectées et transformées. L'API devra comporter au minimum des endpoints de consultation permettant d'interroger les dessertes ferroviaires selon différents critères (ex. : ville de départ, ville d'arrivée, type de train). Une documentation technique claire, incluant des exemples de requêtes, devra être fournie pour garantir la prise en main par des utilisateurs tiers.

### 5. Documentation technique complète

Un rapport technique devra être remis. Il devra détailler :

- les sources de données retenues et leur justification,
- les choix techniques opérés (langages, librairies, frameworks),
- les étapes du processus ETL (collecte, nettoyage, transformation, stockage),
- le schéma de la base de données,
- les spécifications et tests réalisés sur l'API,
- les éventuelles limites identifiées et les pistes d'amélioration.

### 6. Tableau de bord de contrôle

ObRail Europe souhaite disposer d'un tableau de bord simple permettant de visualiser la complétude des données (répartition des trajets jour/nuit, volume de données collectées par opérateur, taux de valeurs manquantes, etc.). Ce tableau de bord devra être accessible à tout public.

### 7. Support de soutien

En complément des livrables techniques, l'équipe projet devra préparer un support de présentation destiné à la soutenance finale devant le client (public technique). Ce support devra synthétiser les principaux éléments du travail réalisé : démarche suivie, difficultés rencontrées, solutions mises en place, résultats obtenus et perspectives.



Il est important de souligner que l'évaluation de cette MSPR repose sur la **combinaison des trois éléments suivants** :

- la qualité du **travail réalisé** au cours du projet,
- la pertinence et l'exhaustivité des **livrables remis** (scripts, modèles, base de données, API, documentation),
- et la capacité de l'équipe à **présenter, justifier et valoriser** ce travail lors de la **soutenance orale**.

**Les équipes devront donc s'assurer que la soutenance reflète bien l'ensemble des compétences attendues**, en démontrant à la fois la maîtrise technique et la capacité à communiquer efficacement auprès d'un client professionnel.

## V – RESSOURCES FOURNIES

Afin de faciliter la réalisation du projet, ObRail Europe met à disposition de l'équipe un ensemble de ressources techniques et documentaires. Ces éléments doivent être utilisés comme points d'appui et ne dispensent pas l'équipe de mener ses propres recherches et choix techniques.

### 1. Jeux de données exemples

ObRail Europe, dans le cadre de ses travaux préparatoires internes, a déjà réalisé un test de constitution d'une base de données unifiée à partir de plusieurs lignes ferroviaires.

L'exemple présenté ci-dessous peut servir de source d'inspiration pour vos propres travaux.

Il ne s'agit que d'une démonstration : vos démarches, vos recherches et vos scripts ETL pourront vous conduire à construire une base différente selon vos choix méthodologiques.

Cet échantillon de données n'est pas exhaustif, mais il illustre les principales structures et formats que vous serez amenés à manipuler.

| trip_id      | agency_name           | route_name                                | train_type               | service_type | origin_stop_name    | destination_stop_name  | departure_time | arrival_time | distance_km | duration_h | emission_gco2e_pkm | total_emission_kgco2e | frequency_per_week | source_dataset               | traction   |
|--------------|-----------------------|---|--------------------------|--------------|---------------------|------------------------|----------------|--------------|-------------|------------|--------------------|-----------------------|--------------------|------------------------------|------------|
| AT_NJ_VIE    | ÖBB Nightjet          | Wien Hbf – Berlin Hbf                     | N Train de nuit intern   | Nuit         | Wien Hbf            | AT Berlin Hbf          | DE 21:00       | 07:05        | 680         | 10.1       | 14.0               | 9.52                  | 7                  | Back-on-Track                | électrique |
| BE_ES_BXL    | European Sleeper      | Bruxelles-Midi – Praha hl.n.              | Train de nuit intern     | Nuit         | Bruxelles-Midi      | BE Praha hl.n.         | CZ 19:00       | 10:30        | 900         | 15.5       | 14.0               | 12.6                  | 3                  | European Sleeper             | électrique |
| CH_IC_LSN    | SBB CFF FFS           | Lausanne – Bern                           | Intercité                | Jour         | Lausanne            | CH Bern                | CH 09:02       | 09:59        | 101         | 0.95       | 6.7                | 0.68                  | 168                | SBB                          | électrique |
| DE_NT_BER    | ÖBB Nightjet          | Berlin Hbf – Paris Est                    | N Train de nuit intern   | Nuit         | Berlin Hbf          | DE Paris Est           | FR 20:18       | 10:24        | 1050        | 14.1       | 14.0               | 14.7                  | 7                  | Nightjet                     | électrique |
| EU_ES_LON    | Eurostar              | London St Pancras – Paris Gare du Nord    | International grand      | Jour         | London St Pancras   | GB Paris Gare du Nord  | FR 14:04       | 17:09        | 373         | 3.1        | 6.0                | 2.24                  | 49                 | Eurostar Tim                 | électrique |
| EU_ES_LON    | Eurostar              | London St Pancras – Paris Gare du Nord    | International grand      | Jour         | London St Pancras   | GB Paris Gare du Nord  | FR 09:31       | 12:47        | 492         | 2.3        | 6.0                | 2.95                  | 49                 | Eurostar Tim                 | électrique |
| FI_NT_HKI    | VR Night Train        | Helsinki – Rovaniemi                      | N Train de nuit          | Nuit         | Helsinki            | FI Rovaniemi           | FI 21:00       | 07:45        | 829         | 10.75      | 14.0               | 11.61                 | 7                  | VR                           | électrique |
| FR_ICN_PAI   | SNCF Intercité        | Paris Austerlitz – Biarritz               | Intercité de nuit        | Nuit         | Paris Austerlitz    | FR Biarritz            | FR 21:30       | 07:05        | 700         | 9.6        | 45.0               | 31.5                  | 7                  | SNCF / Back-on-Track         | mixte      |
| FR_IC_CLF    | SNCF Intercité        | Clermont-Ferrand – Paris                  | Intercité                | Jour         | Clermont-Ferrand    | FR Paris Bercy         | FR 06:28       | 09:42        | 420         | 3.2        | 11.8               | 4.96                  | 49                 | SNCF                         | électrique |
| FR_TER_BOI   | SNCF TER              | Bordeaux – Arcachon                       | Régional                 | Jour         | Bordeaux St-Jean    | FR Arcachon            | FR 10:05       | 10:55        | 66          | 0.83       | 29.2               | 1.93                  | 140                | transport.data.gouv.fr       | électrique |
| FR_TER_GRI   | SNCF TER              | Grenoble – Gap                            | Régional                 | Jour         | Grenoble            | FR Gap                 | FR 12:10       | 14:40        | 100         | 2.5        | 90.0               | 9.0                   | 35                 | transport.data.gouv.fr       | diesel     |
| FR_TGV_PAI   | SNCF TGV              | Paris Gare de Lyon – Bordeaux-Mérignac    | Grande vitesse           | Jour         | Paris Gare de Lyon  | FR Bordeaux-Mérignac   | FR 08:15       | 12:30        | 680         | 4.3        | 3.2                | 2.18                  | 4                  | Rail Europe                  | électrique |
| IT_ICN_MIL   | Trenitalia Intercité  | Milano Centrale – Lecce                   | Intercité de nuit        | Nuit         | Milano Centrale     | IT Lecce               | IT 21:50       | 07:20        | 1000        | 9.5        | 14.0               | 14.0                  | 7                  | Trenitalia                   | électrique |
| IT_ICN_ROM   | Trenitalia Intercité  | Roma Termini – Palermo                    | Intercité de nuit        | Nuit         | Roma Termini        | IT Palermo Centrale    | IT 23:00       | 11:57        | 920         | 13.0       | 14.0               | 12.88                 | 7                  | Back-on-Track                | électrique |
| LT_REG_VNI   | LTG Link              | Vilnius – Klaipėda                        | Régional                 | Jour         | Vilnius             | LT Klaipėda            | LT 08:00       | 12:00        | 320         | 4.0        | 90.0               | 28.8                  | 14                 | LTG Link                     | diesel     |
| RS_ME_ICN    | ŽPČG / Srbija         | Beograd Centar – Bar                      | Intercité de nuit intern | Nuit         | Beograd Centar      | RS Bar                 | ME 20:10       | 08:15        | 476         | 12.1       | 45.0               | 21.42                 | 7                  | ŽPČG / Srbija                | mixte      |
| UK_CS_LON    | Caledonian Sleeper    | London Euston – Edinburgh Waverley        | Train de nuit            | Nuit         | London Euston       | GB Edinburgh Waverley  | GB 23:50       | 07:20        | 634         | 7.5        | 45.0               | 28.53                 | 7                  | Caledonian Sleeper           | mixte      |
| EU_LUX_PAR   | Belmond               | Paris Gare de l'Est – Venezia Santa Lucia | Train de luxe intern     | Nuit         | Paris Gare de l'Est | IT Venezia Santa Lucia | IT 19:45       | 09:35        | 1350        | 13.8       | 25                 | 33.75                 | 2                  | Belmond / Caledonian Sleeper | électrique |
| EU_HISTORH   | Trans-Europe Express  | Amsterdam Centraal – Zermatt              | Intercité historique     | Jour         | Amsterdam Centraal  | NL Milano Centrale     | IT 08:00       | 18:00        | 1000        | 10.0       | 20                 | 20.00                 | 1                  | Archives UIC                 | électrique |
| FI_FICTIF_RC | Polar Express         | Rovaniemi – Pöle Nord                     | Train touristique        | Nuit         | Rovaniemi           | FI Pöle Nord           | AQ 22:00       | 06:00        | 1200        | 8.0        | 45                 | 54.00                 | 1                  | Fictif                       | mixte      |
| UK_NT_LON    | Great Western Railway | London Paddington – Penzance              | Train de nuit            | Nuit         | London Paddington   | GB Penzance            | GB 23:45       | 07:53        | 515         | 8.13       | 90                 | 46.35                 | 7                  | GWR / Back-on-Track          | diesel     |
| CH_SCN_ZEI   | MGB / RhB             | Zermatt – St. Moritz                      | Train panoramique        | Jour         | Zermatt             | CH St. Moritz          | CH 08:52       | 17:35        | 291         | 8.7        | 6.7                | 1.95                  | 7                  | SBB / Glacier Express        | électrique |

### 2. Accès à des sources de données ouvertes

ObRail fournit une liste de sources de données open data pertinentes (par exemple : Back-on-Track Night Train Database, plateformes nationales comme transport.data.gouv.fr, ou bases européennes telles qu'Eurostat). Ces références doivent servir de point de départ à la recherche et à la justification des données intégrées dans le projet.

- European Data Portal – Portail officiel des données ouvertes européennes.
- Eurostat – Statistiques européennes, notamment sur les transports et l'énergie.
- [Transport.data.gouv.fr](https://transport.data.gouv.fr) – Portail français open data dédié aux transports.
- OpenMobilityData (ex TransitFeeds) – Jeux de données GTFS (transport public international).  
<https://mobilitydatabase.org/>
- Transitland : <https://www.transit.land/>
- Husahuc : [https://gitlab.com/husahuc/gtfs\\_france](https://gitlab.com/husahuc/gtfs_france)
- Back-on-Track Night Train Database – Données sur les trains de nuit en Europe.

### 3. Environnement et outils recommandés

- Langages et bibliothèques : Python (pandas, requests, SQLAlchemy), R (tidyverse), SQL.
- Gestion des environnements : Docker pour la conteneurisation et Postman pour le test des APIs.
- Base de données cible : PostgreSQL (ou équivalent compatible).
- Frameworks possibles pour l'API : Flask, FastAPI, Node.js (libre choix sous réserve de documentation claire).

### 4. Assistance et périmètre

Dans le cadre de ce projet pédagogique, l'équipe projet n'aura aucun contact direct avec ObRail Europe. Le cahier des charges constitue la seule expression officielle du besoin. Toute demande de clarification devra être traitée avec l'encadrant pédagogique, jouant le rôle du client.

### 5. Webographie

- PostgreSQL Documentation – Référence officielle du SGBD recommandé :  
<https://www.postgresql.org/docs/>
- Pandas Documentation – Manipulation et nettoyage des données avec Python :  
<https://pandas.pydata.org/docs/>
- FastAPI Documentation – Créer une API REST performante en Python : <https://fastapi.tiangolo.com/>
- Postman – Outil de test et documentation des APIs : <https://www.postman.com/>
- CNIL – RGPD et bases de données – Bonnes pratiques légales et conformité : <https://www.cnil.fr/fr/rgpd-par-ou-commencer>