

Big Data for Labour Market Intelligence

Programme de développement des capacités 2022

Module 1 : Formation technique

Session 5

Accès à la base de données, au modèle de données et aux principales caractéristiques - Conseils pratiques

Systeme de production des donnees

Tunisie (avril 2020-juillet 2022)

- ✓ 680 191 OEL - > 175 203 dedupliques

Ukraine - general (avril 2020-juillet 2022)

- ✓ 2 571 655 OEL - > 1 304 262 dedupliques

Georgie (avril 2021-juillet 2022)

- ✓ 129 271 OEL - 84 817 dedupliques

Egypte (nouveau)

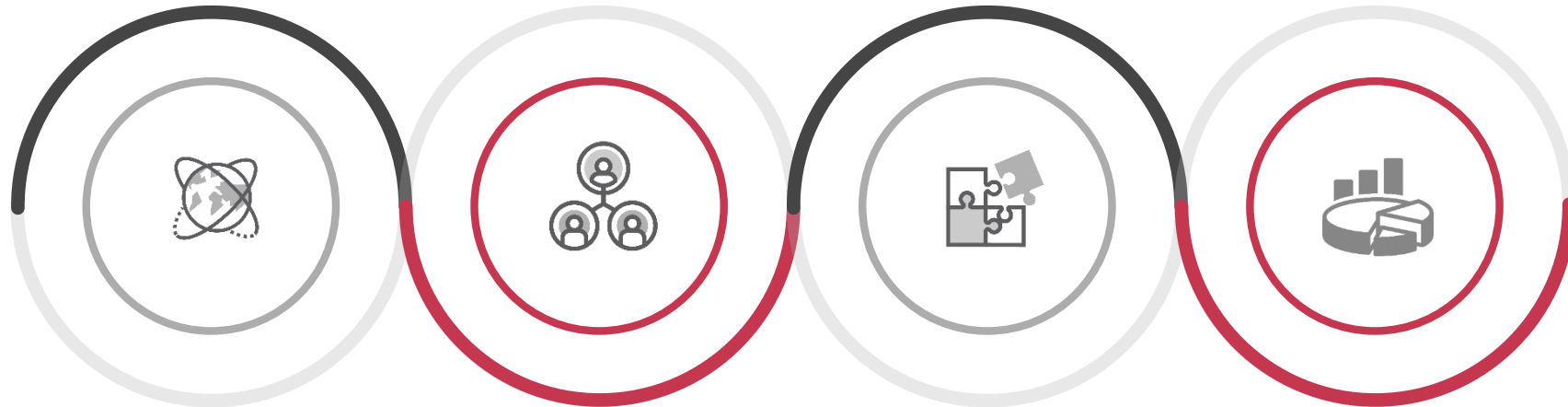
- ✓ 1 307 678 OEL - 391 701 dedupliques

Kenya (nouveau)

- ✓ (la collecte a commence en septembre 2022)

Collecte et décodage des données sur le marché du travail

Les données sur le marché de l'emploi en temps réel fournissent des informations actualisées que les sources traditionnelles ne permettent pas d'obtenir.



Saisie de données sur le marché du travail

Marquage et structuration

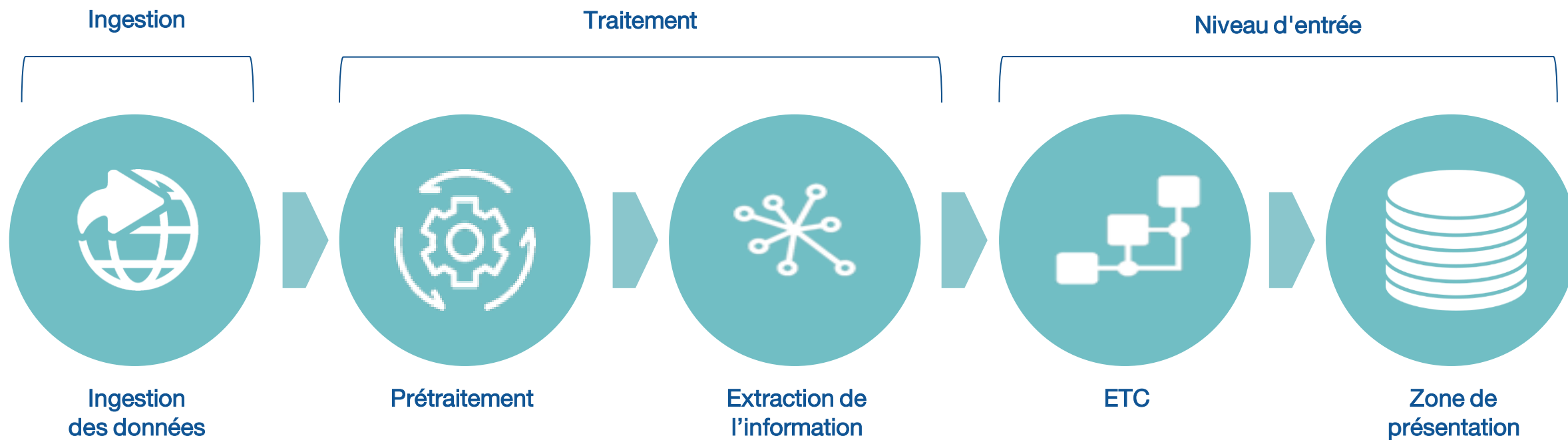
Langue commune

Formulation de conclusions

L'ontologie des données permet de faire des comparaisons

Indication sur les compétences recherchées et les tendances professionnelles dans la vie réelle

Flux de données global



Ingestion des données

- **Objectif :**
 - Collecte d'une quantité considérable de données hétérogènes provenant d'un nombre imprévisible de sources.
- **Défis :**
 - Garantir l'exhaustivité et la cohérence des données.
- **Approche :**
 - Développer un cadre multitechnique (crawling, scraping, API) pour faire correspondre les différentes caractéristiques des sites Web : volume d'offres d'emploi, langues utilisées, structure technologique, approche non invasive et accords politiques.
 - Prévenir les pertes de sources de données par le biais de politiques de redondance.
 - Détecter et collecter les métadonnées pour améliorer la valeur des informations.
- **Caractéristiques :**
 - Pour garantir la gouvernance du processus dans la phase d'ingestion des données, nous avons développé un système de surveillance avec un outil de planification et un module d'alerte.

Prétraitement des données

- **Objectif**
 - Alimenter la phase d'extraction des informations avec les bonnes données.
- **Défis**
 - Mesurer, surveiller et améliorer la qualité des données, pour en maximiser l'exhaustivité, la cohérence, la complexité, l'actualité et la périodicité.
- **Approche**
 - Développer un pipeline à plusieurs phases, axé sur :
 - la détection des offres d'emploi : analyse d'une page Web pour sélectionner uniquement le contenu relatif aux offres d'emploi.
 - la déduplication : détecter les offres d'emploi doubles pour obtenir une seule entité d'offre d'emploi.
 - la détection des dates : identifier les dates de publication et d'expiration grâce à l'analyse de la description des offres d'emploi.
 - la durée de l'offre : méthode pour définir la date d'expiration, lorsqu'elle n'est pas explicitement disponible.
- **Caractéristiques** :
 - Garantir la qualité des données pendant toutes les phases du traitement.

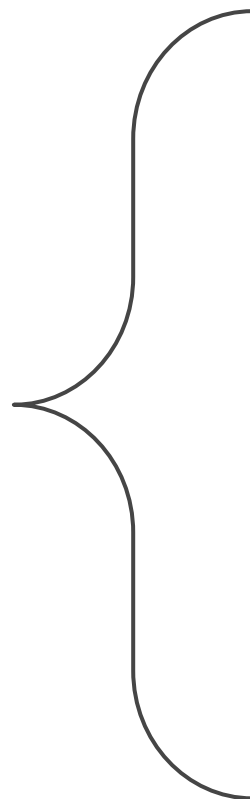
Classification des données

- **Objectif :**
 - Extraire et structurer les informations des données à fournir au niveau présentation.
- **Défis :**
 - Traiter des quantités considérables de données hétérogènes écrites dans des langues différentes.
- **Approche :**
 - Développer un cadre adaptable, dépendant de la langue, adapté aux différentes caractéristiques des informations. Quelques défis pertinents :
 - Classification des caractéristiques de Profession : méthodes combinées comme l'apprentissage automatique, la modélisation thématique et l'apprentissage non supervisé.
 - Classification des caractéristiques de Compétences : autres méthodes combinées, comme l'analyse de texte avec similarité du corpus ou des connaissances.
- **Caractéristiques :**
 - Garantir l'extraction d'informations explicables, les méthodes de classification des collectes et les caractéristiques pertinentes.

Espace Présentation



Zone de
présentation



Capacités de
découverte des
données

Publier, partager
et collaborer

Analyse et BI en
libre-service

Analyse avancée
intégrée

Enquête visuelle

Systèmes
d'information

Intelligence
géospatiale

Découverte des
graphiques et
enquête

Data
storytelling (mise
en récit des
données)

Identifier les utilisateurs et les modèles de navigation



Analystes de
données
et experts en
métadonnées



Décideurs et
utilisateurs
professionnels

Experts en mégadonnées et analystes de données



Capacités de découverte des données



Publier, partager et collaborer



Intégration de l'apprentissage automatique



Analyse avancée intégrée

Décideurs et utilisateurs professionnels



Analyse et BI en
libre-service



Enquête visuelle

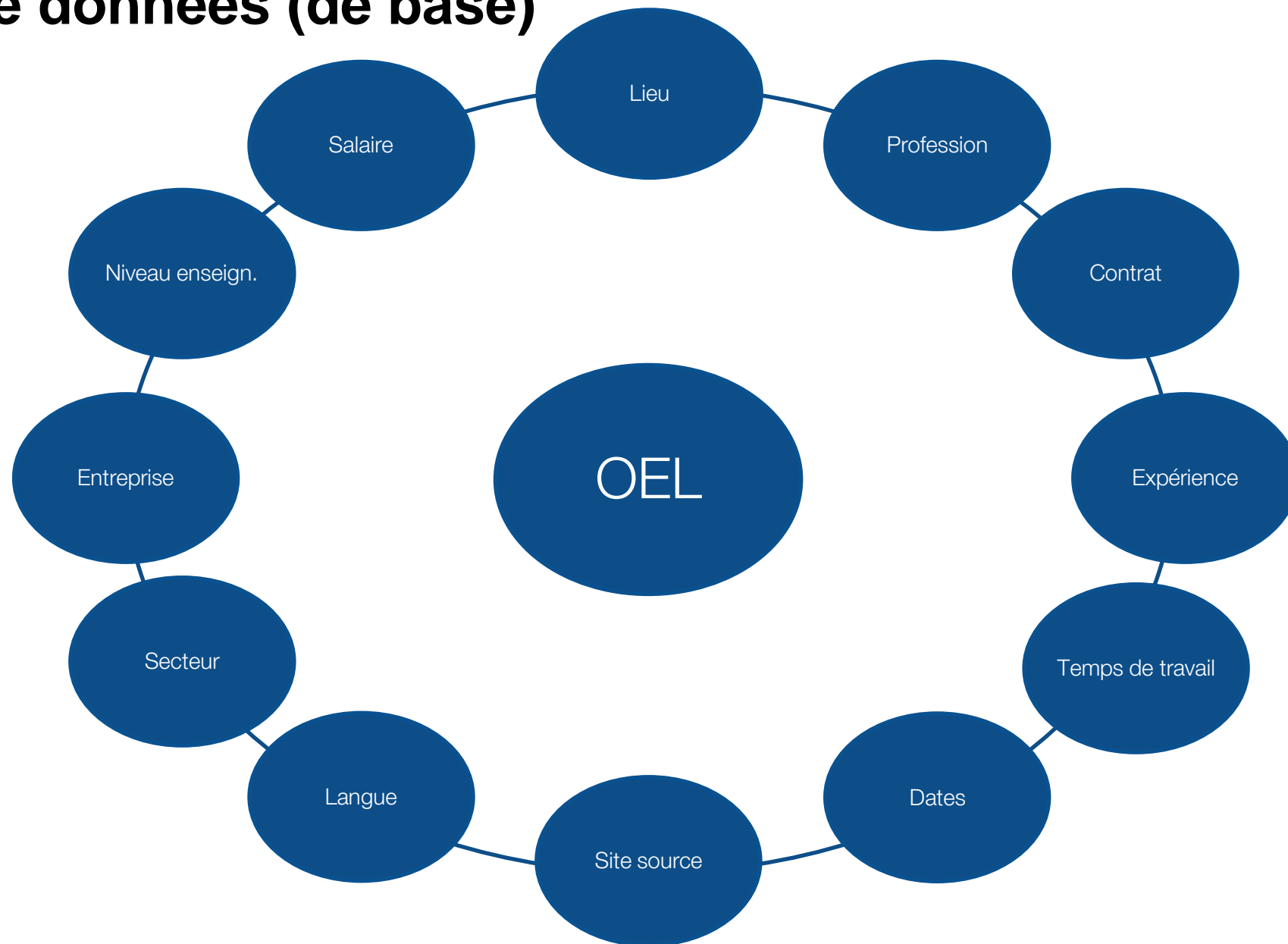


Découverte
gouvernée des
données



Data
storytelling (mise en
récit des données)

Modèle de données (de base)



Secteur

Référence du poste : 990-NHSE8576N
Industrie : Santé
Salaire : 56 665 - 69 168 par an
Lieu : Leeds

NHS England dirige le service national de santé (NHS) en Angleterre. Nous fixons les priorités et l'orientation du NHS, et nous encourageons et informons le débat national pour améliorer la santé et les soins. Nous voulons que chacun et chacune ait un meilleur contrôle sur sa santé et sur son bien-être, et qu'il ou elle puisse vivre plus longtemps et en meilleure santé grâce à des services de santé et de soins de haute qualité, empathiques, inclusifs et en constante amélioration...

La norme NACE était présente en tant que champ structuré dans l'OEL (évaluée comme « Santé », ce qui correspond à l'ontologie).



Valeur de classification :
86 - Activités pour la santé humaine

Référence



Contrat

Ouvrier agricole - Aylesbury
Contrat : temporaire (3 mois)
Salaire : 10 £ par heure

Nous sommes actuellement à la recherche d'un travailleur persévérant et honnête dans la région d'Aylesbury.
Vous serez le principal ouvrier agricole du site et vous devrez notamment nettoyer le site et aider les autres personnes présentes sur le site. Vous pouvez rejoindre le site par les transports publics ou en voiture si vous...

Le contrat était présent en tant que champ structuré dans l'OEL (évalué comme « temporaire », ce qui correspond à la valeur ontologique).



Valeur de classification :
Temporaire

Référence

Temps plein
Emploi indépendant
Temporaire

Taxonomie
personnalisée

Temps de travail

Référence du poste : 184-SS.GEN.38

Département : sénilité

Lieu : Bracken House, Chard

L'équipe de santé mentale communautaire pour les personnes âgées de Chard recherche activement un infirmier (H/F) en santé mentale communautaire de classe 5 **à temps partiel** pour aider le service de consultation mémoire et l'hôpital de jour.

Au sein d'une équipe intégrée innovante, vous travaillerez en étroite collaboration avec le service des soins infirmiers, l'équipe de rééducation intégrée et l'équipe médicale, ainsi qu'avec les médecins généralistes, les services sociaux pour adultes, le service des soins de courte durée et le secteur du volontariat...

Les heures de travail n'étaient pas présentes en tant que champ structuré dans l'OEL, mais le texte contient une référence au temps de travail (« temps partiel ») qui correspond à l'ontologie.



Valeur de classification :
Temps partiel

Référence

Temps plein
Temps partiel

Taxonomie
personnalisée

Niveau d'enseignement

Rôle : Chef d'équipe du matériel roulant

Lieu : sud de Londres

Salaire : environ 47 500 £

Niveau d'études : deux ans d'études supérieures

Expérience : moins d'un an

L'objectif est de diriger les activités quotidiennes pour assurer la livraison des stocks en temps voulu tout en veillant au respect des normes d'entretien concernées et au bon déroulement des activités dans l'environnement des passagers.

Les rôles et les responsabilités comprennent, sans s'y limiter : livraison quotidienne de la flotte mise en service de manière fiable et cohérente, membre de l'équipe de direction qui fournit un entretien rentable et efficace, ...

Le niveau d'enseignement était présent en tant que champ structuré dans l'OEL sous l'expression « Niveau d'études » (évalué comme « deux ans d'études supérieures », ce qui correspond à un titre alternatif de l'ontologie).



Valeur de classification :
Baccalauréat ou équivalent

Référence

CITE
2011

Salaire

Rôle : Chef d'équipe du matériel roulant

Lieu : sud de Londres

Salaire : environ 47 500 £

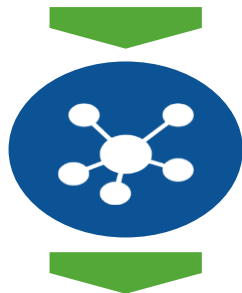
Niveau d'études : deux ans d'études supérieures

Expérience : moins d'un an

L'objectif est de diriger les activités quotidiennes pour assurer la livraison des stocks en temps voulu tout en veillant au respect des normes d'entretien concernées et au bon déroulement des activités dans l'environnement des passagers.

Les rôles et les responsabilités comprennent, sans s'y limiter : livraison quotidienne de la flotte mise en service de manière fiable et cohérente, membre de l'équipe de direction qui fournit un entretien rentable et efficace, ...

Le salaire était présent en tant que champ structuré dans l'OEL (évalué à « 47 500 £ par an » et converti en euros).



Valeur de classification :
48 000 - 54 000 € par an

Référence

13 niveaux

Taxonomie
personnalisée

Expérience

Rôle : Chef d'équipe du matériel roulant

Lieu : sud de Londres

Salaire : environ 47 500 £

Niveau d'études : deux ans d'études supérieures

Expérience : moins d'un an

L'objectif est de diriger les activités quotidiennes pour assurer la livraison des stocks en temps voulu tout en veillant au respect des normes d'entretien concernées et au bon déroulement des activités dans l'environnement des passagers.

Les rôles et les responsabilités comprennent, sans s'y limiter : livraison quotidienne de la flotte mise en service de manière fiable et cohérente, membre de l'équipe de direction qui fournit un entretien rentable et efficace, ...

L'expérience était présente en tant que champ structuré dans l'OEL (évaluée comme « moins d'un an », ce qui correspond à l'ontologie).



Valeur de classification :
Jusqu'à un an

Référence

8 niveaux

Taxonomie
personnalisée

Lieu

Référence du poste : 990-NHSE8576N
Industrie : Santé
Salaire : 56 665 - 69 168 par an
Lieu : Leeds

NHS England dirige le service national de santé (NHS) en Angleterre. Nous fixons les priorités et l'orientation du NHS, et nous encourageons et informons le débat national pour améliorer la santé et les soins. Nous voulons que chacun et chacune ait un meilleur contrôle sur sa santé et sur son bien-être, et qu'il ou elle puisse vivre plus longtemps et en meilleure santé grâce à des services de santé et de soins de haute qualité, empathiques, inclusifs et en constante amélioration...

Le lieu était présent en tant que champ structuré dans l'OEL (évalué comme « Leeds », ce qui correspond à l'ontologie).



Valeur de classification :
Leeds

Référence

NUTS/ISO
3166
ET
UAL

Profession

Technicien Unix

Dans ce rôle, vous serez responsable des activités suivantes :

- o Installation et support du système d'exploitation du serveur, du logiciel de gestion du système et des utilitaires du système d'exploitation.
- o Gestion de la configuration du système d'exploitation
- o Gestion des systèmes de fichiers et les files d'attente d'impression
- o Surveillance et conservation les fichiers journaux du système d'exploitation
- o Recommandation de mises à jour du système d'exploitation et des modifications de la configuration, ...

L'algorithme d'apprentissage automatique a trouvé la bonne profession, non présente dans l'ontologie



Valeur de classification :
2522 - Administrateurs de systèmes

Référence
(ESCO 4^e niveau)

ESCO v1.1
CITP-08 (4
CHIFFRES)

Compétences

Vous êtes un administrateur expérimenté à la recherche de votre prochain contrat dans la région de Bristol ?

Mon client est une grande entreprise spécialisée dans l'entretien de biens immobiliers et cherche un administrateur de succursale (H/F) à intégrer immédiatement dans son équipe en commençant par une mission intérimaire.

Le(la) candidat(e) sélectionné(ée) devra réaliser de nombreuses tâches **administratives**, y compris **répondre aux appels entrants, collaborer avec les sous-traitants et émettre des factures**.

Responsabilités :

- **Utiliser les systèmes informatiques** pour prêter un service administratif dans la préparation, le traitement et la sélection des devis, des offres et des appels d'offres.
- **Commander** des biens, du matériel et des services pour répondre aux exigences des contrats.
- **Gérer les communications interne et externe** et enregistrer et/ou rapporter les informations le cas échéant.
- S'assurer que toutes les données contractuelles, la documentation et les rapports nécessaires sont exacts et produits en temps voulu.
- Apporter son soutien à la direction afin de **répondre aux besoins de l'entreprise**.
- **Gérer les requêtes et/ou communications des clients** de manière professionnelle et efficace.

Profil souhaité :

- Compétences informatiques solides, **maîtrise de MS Office**
- Excellentes **capacités de communication** (écrit et oral)
- Excellente organisation avec **compétences de gestion du temps** avérées
- ...

Référence

ESCO V1.1
+
personnalisation

Jeu de données



Jeu de données – Mots-clés pour le modèle et le magasin de données

- Tableaux dénormalisés
- Magasin en colonnes

Modèle de jeu de données

2 Tableaux

Document FT

- 1 ligne pour chaque :
- General_ID [?] clé
 - OEL
 - Source
 - Lieu

Pourquoi ? Parce que, pour chaque OEL, nous pouvons détecter un poste à plusieurs endroits

par ex. : « Concepteur de logiciels à Londres / Liverpool »

Analyse des compétences FT

- 1 ligne pour chaque :
- General_ID [?] clé
 - OEL
 - Source
 - Lieu
 - Compétence

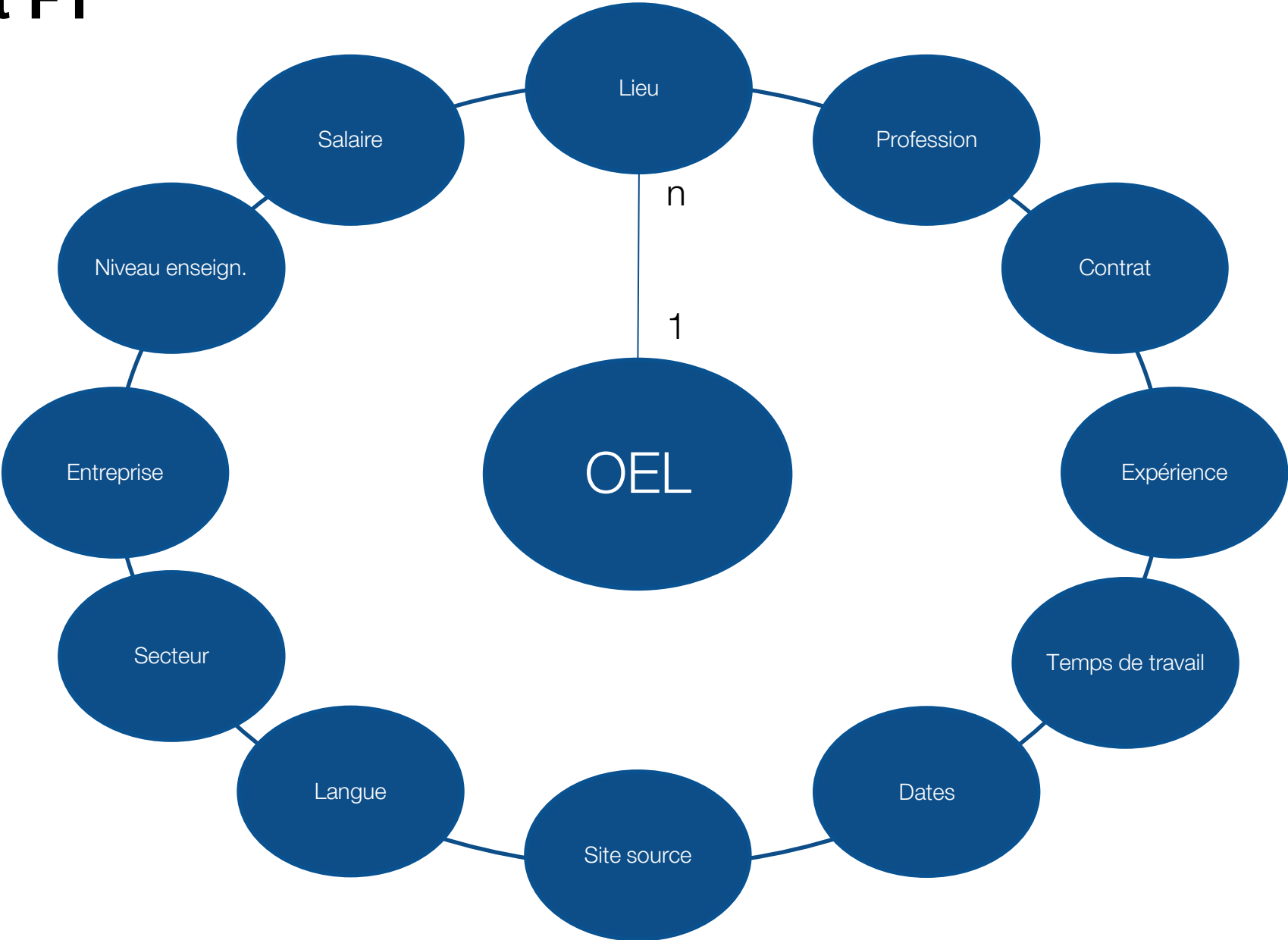
Pourquoi ? Parce que, pour chaque OEL, nous pouvons clairement détecter plusieurs compétences

par ex. : « Concepteur de logiciel à Londres / Liverpool, avec une culture en matière d'orientation client, qui parle anglais et résiste au stress. »

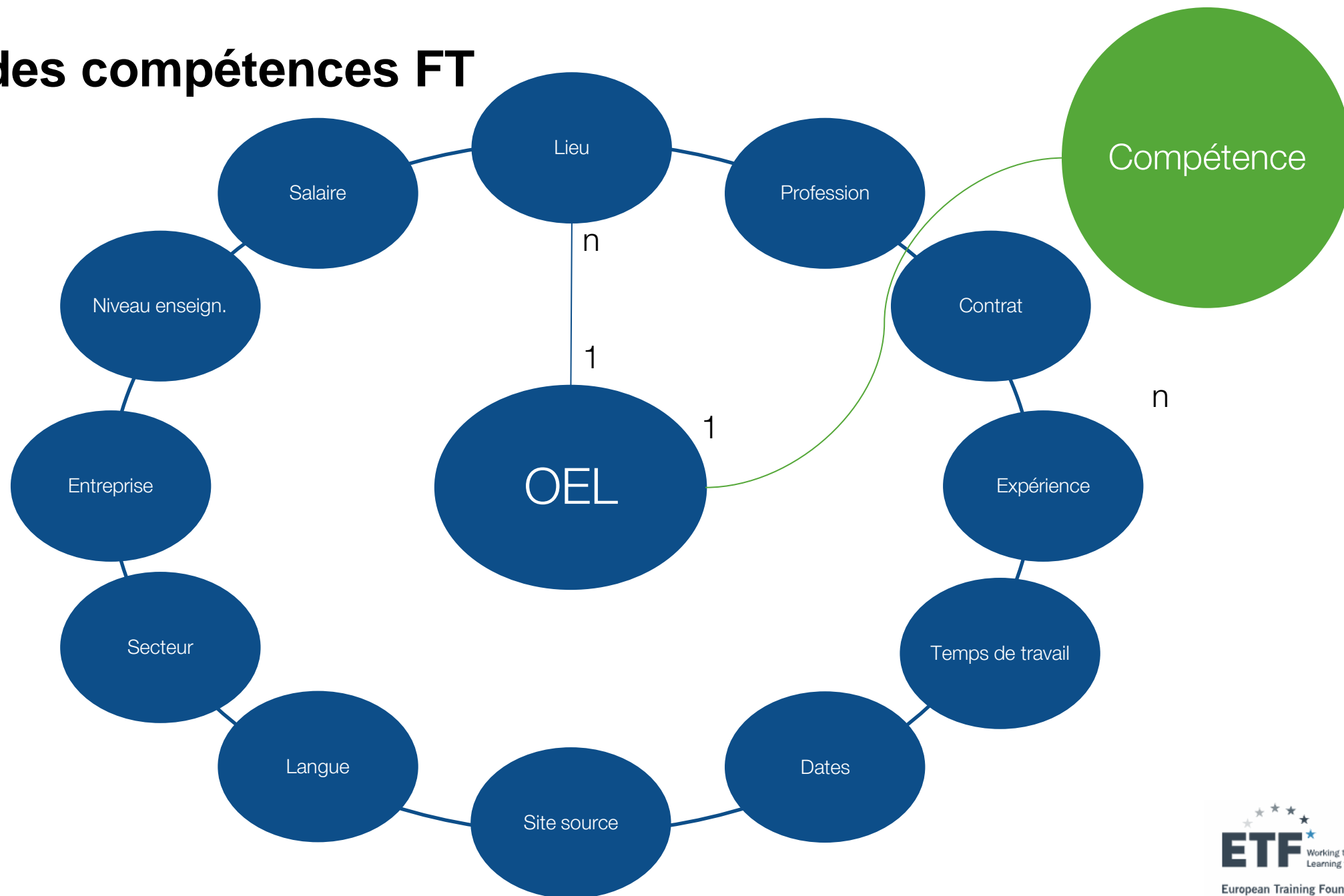
Modèle de jeu de données

Ainsi, pour obtenir le nombre d'offres d'emploi, il faut toujours calculer un compte unique par General_ID

Document FT



Analyse des compétences FT



Outils



Amazon EMR



Amazon Athena

