

**Big Data pour l'information sur le marché du travail :  
Analyse des offres d'emploi en ligne (OJV)**

Programme de formation virtuelle

Programme complet : 8, 10 et 15 juin 2021

Multilingue : anglais-français-russe

Inscription : [lien](#)

**Ordre du jour :**

**Jour 1 : 08/06/2021, 09 h 00-14 h 30 CET (GMT+2)**


- IMT - basée sur le données Web : des offres d'emploi en ligne aux données et analyses des compétences, des professions et des secteurs. Nouvelles données, nouvelles analyses. Aperçu, méthodologie, utilisations.
- Étude de cas

**Jour 2 : 10/06/2021, 09 h 00-14 h 30 CET (GMT+2)**



- IMT – basées sur les données Web : accent mis sur l'ingestion, le traitement et la classification des données, avec des processus assistés par l'IA.
- Étude de cas

**Jour 3 : 15/06/2021, 09 h 00-14 h 00 CET (GMT+2)**

- IMT – basée sur les données web : cas d'utilisation, exercices pratiques et leçons apprises



ETF travaille avec Burning Glass Europe

[Link to access](#)  
[Link of the Webpage](#)

### Projet ETF Big Data pour le Système d'information sur le marché du travail (SIMT)

Démarré au second semestre 2018, le projet d'innovation de l'ETF a d'abord élaboré un bref manuel méthodologique « [Big Data for labour market intelligence : an introductory guide](#) » (Big Data pour le Système d'information sur le marché du travail : un guide d'introduction) (2019). Un [programme de formation](#) pour les experts du réseau « Make it Match » du Partenariat oriental a été couronné de succès en novembre 2019. En 2020, le système de données complet pour l'analyse des OEL a été établi, avec des tableaux de bord de données disponibles ici: [Tunisie](#) et [Ukraine](#). Les tableaux de bord sont mis à jour avec les données des offres d'emploi en ligne collectées jusqu'au 31 mars 2021. Le système de données est axé sur la demande (offres d'emploi en ligne - OEL).

Le projet Big Data pour SIMT de l'ETF se poursuit en 2021, avec l'ajout d'un nouveau pays et la poursuite de l'ingestion, de la classification et de l'analyse des données sur les offres d'emploi en ligne (OEL) pour la Tunisie et l'Ukraine. De nombreuses nouvelles questions et requêtes peuvent être analysées sur la base de cette base de données croissante d'OEL, comme le prévoit le paradigme « laisser parler les données ». De nouvelles idées seront identifiées à partir de ce vaste ensemble de données, et un « tableau de bord vert » sera développé.

**Ce programme de formation virtuelle de trois jours est conçu pour répondre à ces questions.**

- Mais qu'est-ce que l'analyse des Big Data pour les systèmes d'information sur le marché du travail ?
- Comment explorer et créer de la valeur à partir de grands volumes d'offres d'emploi en ligne pour un SIMT en temps réel ?
- Quelles dimensions et questions liées aux compétences, aux professions et à la dynamique du marché du travail peuvent être analysées à l'aide du Big Data / SIMT en temps réel ?

- Quelles sont les exigences des pays partenaires de l'ETF pour exploiter ces nouvelles sources de données et ces nouveaux systèmes ?
- Comment le SIMT en temps réel basé sur le Big Data et la classification assistée par l'IA peuvent-ils être utilisés et compléter les statistiques nationales établies ?

**Participants** : représentants des départements de recherche et d'analyse des ministères (éducation, formation et travail) et des centres de recherche spécialisés de tous les pays partenaires de l'ETF. La participation à toutes les activités prévues dans le programme de trois jours est recommandée (obligatoire) pour tous les participants inscrits.

Cette formation virtuelle de trois jours mobilise les participants 4,5 heures par jour et intercale des journées de révision et de préparation des blocs thématiques du programme de formation.

### Calendrier du programme de formation

**Jour 1 : 08/06/2021: 09 h 00 – 14 h 30 (CET) – 07 h 00-12 h 30 GMT**

Heure (CET)	Session	Orateur
09 00-09 h 30	Ouverture, accueil des participants. Objectifs de l'atelier de formation. Présentation des participants	Eduarda Castel-Branco, ETF
09 h 30-09 h 45	Bref aperçu du projet Big Data pour SIMT de l'ETF	Eduarda Castel-Branco
09 h 45-11 h 15 Session 1	<b>Big Data pour le Système d'information sur le marché du travail (SIMT)</b> Présentation du système et résultats <ul style="list-style-type: none"> <li>• Aperçu</li> <li>• Défis</li> <li>• Aperçu du flux de données</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
11 h 15-11 h 45	Sondage. Questions et réponses. Discussion	Eduarda Castel-Branco Alessandro Vaccarino, Mauro Pelucchi
11 h 45-12 h 00	<b>Pause-café</b>	
12 h 00-13 h 30 Session 2	<b>Big Data pour le SIMT</b> Description du système <ul style="list-style-type: none"> <li>• Ingestion</li> <li>• Pré-traitement</li> <li>• Classification</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
13 h 30-13 h 45	Questions et réponses. Discussion	Alessandro Vaccarino Mauro Pelucchi
13 h 45-14 h 30	Cas d'étude : Eurostat – Statistiques intelligentes. Zoom sur les données issues d'offres d'emploi en ligne	Eurostat

**Jour 2 : 10/06/2021 : 09 h 00 – 14 h 30 (CET) – 07 h 00-12 h 30 GMT**

Heure (CET)	Session	Orateur
09.00-09.15	Ouverture. Impressions sur le Jour 1.	Eduarda Castel-Branco, ETF

09 h 15- 11 h 00 Session 1	<b>Big Data pour le Système d'information sur le marché du travail (SIMT)</b> Composantes du système <ul style="list-style-type: none"> <li>• Résumé et application pratique (Ingestion et traitement des données)</li> <li>• Show-cooking : extraction de données d'un site d'entreprise</li> <li>• Questions</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
11 h 00- 11 h 30	Questions, clarifications. Sondage.	
11 h 30- 12 h 00	<b>Pause-café</b>	
12 h 00- 13 h 30 Session 2	<b>Big Data pour le SIMT</b> Lab <ul style="list-style-type: none"> <li>• Show-cooking: créer une pipeline de classification</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
13 h 00- 13 h 15	Questions and answers. Discussion	Alessandro Vaccarino Mauro Pelucchi
13 h 45- 14 h 30	<b>Étude de cas :</b> Application d'IMT en temps réel : adjacence des compétences, progression des carrières et choix éclairés.	Orateur : Davor Miskulin, Burning Glass Technologies

**Jour 3 : 15/06/2021 : 09 h 00 – 14 h 15 (CET) – 07 h 00-12 h 15 GMT**

**Exigences techniques : Tableau Public installé sur le poste de travail ([télécharger le lien](#))**

Heure (CET)	Session	Orateur
09 h 00- 09 h 15	Ouverture. Résumé des jours 1 et 2.	Eduarda Castel-Branco, ETF
09 h 15- 11 h 00 Session 1 (EN)	<b>Big Data pour le Système d'information sur le marché du travail (SIMT)</b> Composantes du système : présentation des données <ul style="list-style-type: none"> <li>• Visualisation des données</li> <li>• Show-cooking : visualiser les ressemblances professionnelles</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
11 h 00- 11 h 15	<b>Pause café</b>	
11 h 15- 12 h 00 Session spéciale	<b>ESCO</b> – la taxonomie (classification) des compétences et métiers, qui supporte la mobilité (travail et études) et les liens avec les certifications et l'emploi.	Dimitrios Pikiou ESCO – Direction Générale Emploi, Affaires Sociales et Inclusion
12 h 00- 13 h 15 Session 2	<b>Big Data pour l'IMT</b> <ul style="list-style-type: none"> <li>• Show-cooking : utiliser Tableau Public et présentation exercice</li> <li>• Pratique : analyser les données sur le marché du travail</li> </ul>	Alessandro Vaccarino Mauro Pelucchi
13 h 15- 14 h 15	Partage des résultats et découvertes de la session 2. Feedback final Clôture de l'atelier de formation	Eduarda Castel-Branco Alessandro Vaccarino Mauro Pelucchi

## Contexte général

### Programme de formation : thèmes et résultats

Jour 1 : 8 juin	Jour 2 : 10 juin	Jour 3 : 15 juin
<p><b><u>Titre</u> :</b> Information sur le marché du travail (IMT) en temps réel : <b>Aperçu, objectifs, avantages et cas d'utilisation.</b></p> <p><b>Type de session :</b> Conférence (présentation)</p> <p><b><u>Étude de cas</u> :</b> session supplémentaire (30 min)</p> <p>Eurostat et les statistiques intelligentes, l'utilisation des données web (avec un accent sur l'OEL) - perspectives, portée et exigences. Potentialités pour d'autres agences statistiques nationales ?</p> <p>Conférencier : EUROSTAT</p>	<p><b><u>Titre</u> :</b> L'IMT en temps réel : accent mis sur l'ingestion, le traitement et la classification des données.</p> <p><b>Type de session :</b> Show cooking</p> <p><b><u>Exigences facultatives</u> :</b> Tableau Public, Google Colab, Databricks Community Edition</p> <p><b><u>Étude de cas</u> :</b> session supplémentaire (30 min)</p> <p>Applications de l'IMT en temps réel : adjacence des compétences, carrières et choix éclairés.</p> <p>Conférencier : Davor Miskulin, Burning Glass Technologies</p>	<p><b><u>Titre</u> :</b> Veille stratégique sur les besoins en compétences : cas d'utilisation, exercices pratiques et leçons apprises</p> <p><b><u>Session spéciale</u> :</b> ESCO – la taxonomie européenne des compétences et des professions soutenant la mobilité.</p> <p><b>Type de session :</b> Atelier</p> <p>Les participants analysent les données afin de comprendre en profondeur le type d'informations fournies par le système IMT basé sur le Big Data.</p> <p><b><u>Description</u> :</b> sur la base d'exemples pratiques, les participants sont invités à effectuer de brèves analyses de données à l'aide de Tableau, ou de leurs outils d'analyse de prédilection.</p> <p>Les leçons apprises et les conclusions sont partagées avec le groupe.</p> <p><b><u>Exigences</u> :</b> Tableau Public</p>
<p><b><u>Résultat</u> :</b> les participants acquièrent une vue d'ensemble complète du système d'IMT basé sur les Big Data (OEL). De l'Extraction de Connaissances à partir des Données (KDD) - aux phases de collecte et de traitement des données.</p>	<p><b><u>Résultat</u> :</b> les participants approfondissent leur compréhension du système d'IMT, en montrant le fonctionnement de ses composantes.</p>	<p><b><u>Résultat</u> :</b> les participants sont capables de :</p> <ul style="list-style-type: none"> <li>- discuter et explorer les exigences du Big Data pour les systèmes d'IMT ;</li> <li>- interpréter et questionner les tableaux de bord de données</li> <li>- proposer des questions pour explorer toutes les potentialités du Big Data pour l'IMT</li> <li>- analyser les caractéristiques et les tendances des compétences et des professions sur la base du Big Data pour l'IMT.</li> </ul>

### Big Data pour les Informations sur le marché du travail (IMT)

#### Un nouveau monde d'analyse des données ...

La veille stratégique sur les besoins en compétences ne suffit pas pour comprendre la direction et l'ampleur de la transformation des tâches, des emplois, des compétences et des qualifications provoquée

par une vague de facteurs de changement, qui a stimulé la numérisation de la plupart des processus dans nos sociétés. De nouvelles analyses de données sont apparues pour faire progresser la veille stratégique sur les besoins en compétences et compléter les statistiques, enquêtes et données administratives classiques.

Les données sont considérées comme le nouveau pétrole. La numérisation des processus, des services, des entreprises et des interactions personnelles et sociales génère une masse croissante de données dans le monde entier. L'objectif principal de l'analyse des Big Data est de créer des connaissances à partir de grands volumes de données, disponibles à une vitesse et une variété élevées.

### ... peut être appliqué pour les informations sur le marché du travail

L'intelligence artificielle (IA) et l'apprentissage automatique modifient non seulement le marché du travail, mais nous offrent également de nouveaux outils d'analyse de la main-d'œuvre. Les offres d'emploi ou les annonces d'emploi sont publiées, rafraîchies, mises à jour en grand nombre sur des sites web de différents types, tailles et couvertures. L'exploration des informations inhérentes à une source de données aussi vaste est devenue un objectif des centres de recherche et des organismes publics dans un certain nombre de pays. Ces vastes sources de données sont essentielles pour comprendre la dynamique et le fonctionnement des marchés du travail en ligne, ainsi que l'évolution des choix de recrutement des employeurs.

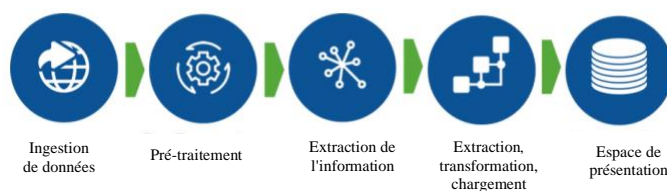
L'analyse du Big Data peut être utilisée pour cartographier les compétences par profession, pour identifier les compétences obsolètes, pour faire une analyse prédictive de la demande de nouvelles professions et compétences, et pour mieux saisir les interactions entre les compétences - en fonction de la granularité des données et quasiment en temps réel.

Dans l'Union européenne, le Cedefop mène depuis 2016 un projet révolutionnaire dans ce domaine et a créé un vaste système de données basé sur l'analyse et la classification de millions d'offres d'emploi en ligne (OEL) des États membres de l'Union européenne (UE). Dans la plateforme [OVATE](#), les résultats sont présentés dans des tableaux de bord interactifs de variables combinées et de différentes couvertures géographiques. En 2021, le Cedefop a renouvelé et mis à niveau les tableaux de bord OVATE et poursuivra l'analyse des relations entre professions, compétences et secteurs. De l'autre côté, Eurostat reprend la partie infrastructure de données, pour se concentrer sur les statistiques et les tendances temporelles détaillées et se focaliser sur la localisation territoriale. Cette nouvelle phase du projet est pilotée conjointement par le Cedefop et Eurostat.

### Big Data pour l'IMT d'ETF

L'application pratique du projet d'ETF a commencé en 2019 avec une analyse de faisabilité des marchés du travail sur le web du Maroc et de la Tunisie, qui a débouché sur un rapport complet évaluant et classant les portails d'**offres d'emploi en ligne** (OEL). La mise en place d'un système intégré de collecte, de traitement, de classification, d'analyse et de visualisation des données a été au cœur des travaux en 2020, dans deux pays pilotes (Ukraine et Tunisie).

### Aperçu schématique du flux de travail et de la méthode du Big Data pour l'IMT

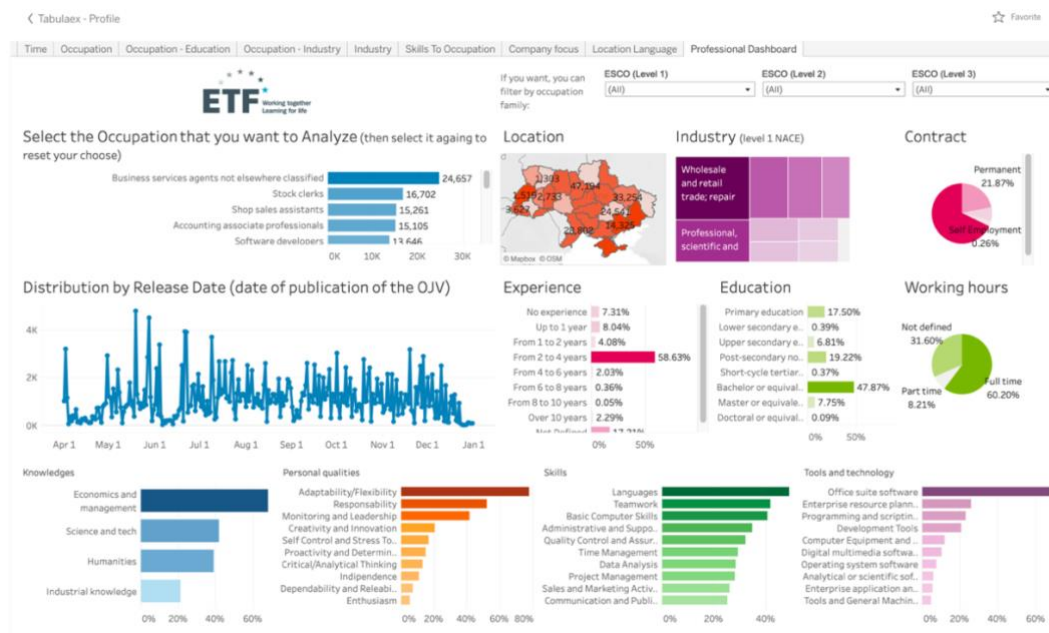


Les principaux résultats des activités de l'ETF en matière de Big Data pour l'IMT, notamment les **tableaux de bord de données, les programmes de formation, le manuel méthodologique et les analyses**, ont été publiés à l'adresse suivante:

- Page web Open Space : [L'IMT en transformation - Gros plan sur le Big Data](#)
- Tableaux de bord de données : [Ukraine](#) et [Tunisie](#)
- Bref manuel méthodologique (anglais) "[Big Data for labour market intelligence : an introductory guide](#)" (2019).
- Bref manuel méthodologique, 2019 ([russe](#))
- Un [programme de formation](#) spécifique pour les analystes et experts en données, novembre 2019.
- Rapport : [Big Data for LMI, Web Labour Market Landscaping Ukraine](#), 2020.
- Rapport : [Big Data for LMI in Ukraine : analytical report](#), 2020.
- [Page web du webinaire sur le Big Data pour le LMIS](#), 10 décembre 2020

En collaboration avec l'équipe de science des données de l'*Université Milano Bicocca* et *Burning Glass Europe* (Italie), la Fondation européenne pour la formation (ETF) a achevé en décembre 2020 une phase décisive de son projet d'innovation « Big Data pour l'Information sur le marché du travail ». Des centaines de milliers d'offres d'emploi en ligne collectées pendant 8 mois consécutifs en 2020 (avril-décembre), traitées et classées automatiquement par rapport à des classifications / taxonomies internationales telles que CITE 2011, NACE, NUTS / ISO et **ESCO** fournissent des informations uniques sur les compétences, les caractéristiques professionnelles et la dynamique des marchés du travail [tunisien](#) et [ukrainien](#). Nous parlons de « perspectives uniques » - en raison de leur granularité et de leur nature instantanée. Certains des nombreux angles d'analyse possibles sont visualisés dans les tableaux de bord de deux pays.

**Figure 1 : Tableau de bord professionnel de l'Ukraine : une vue multidimensionnelle des variables en un instantané**



Le système de données utilise l'ESCO comme référence pour la classification automatique des compétences identifiées dans les centaines de milliers d'OEL. Pour le cas des données OEL tunisiennes, nous avons utilisé l'ESCO - en versions française et anglaise. Mais pour le cas de l'Ukraine, une étape supplémentaire était indispensable : la traduction des **compétences ESCO** en langues **russe et ukrainienne** (plus de 4 000 termes).

L'avantage particulier des OEL en tant que sources d'IMT réside dans le fait qu'elles expriment / représentent la détermination par les employeurs des profils nécessaires aux besoins de l'entreprise ou de l'activité dans une période donnée. Le traitement et la classification automatiques des termes et descriptions de compétences propres aux employeurs montrent des cas de compétences OEL sans correspondance directe avec l'ESCO. Pas de surprise : la transformation technologique et numérique du travail et des compétences est beaucoup plus rapide que le rythme d'alignement et de mise à jour des ESCO. Que faire alors ? L'équipe de science des données impliquée dans le projet a appliqué des techniques d'apprentissage automatique (par exemple, Word2Vec) pour améliorer les compétences ESCO, en créant une correspondance entre une nouvelle compétence « non-ESCO » et une compétence ESCO proche (approximative). Ce processus et la correspondance proposée par la machine sont discutés et validés par des professionnels (humains) du secteur, de la profession et de la technologie concernés. Cette technique et cette approche peuvent-elles avoir une application plus large dans le contexte des mises à jour ESCO ?

27/04/2021. Contact : Eduarda Castel-Branco - [ecb@etf.europa.eu](mailto:ecb@etf.europa.eu)

## Exigences techniques

# Procédure d'accès à Databricks

Suivez les étapes ci-dessous pour créer un compte gratuit et accéder à Databricks *Community Edition*:

1. Rendez-vous sur <https://databricks.com/try-databricks>
2. Remplissez le formulaire. Saisissez une adresse e-mail valide et le lien d'activation vous sera envoyé

## Try Databricks

An open and unified data analytics platform for data engineering, machine learning, and analytics

From the original creators of Apache Spark™, Delta Lake, MLflow, and Koalas

Tell us a little about yourself to get started.

* First Name: Mauro	* Last Name: Pelucchi
* Company Name: Unimib	* Work Email: mauro.pelucchi@unimib.it
* How would you describe your role?: Data Scientist	* What is your intended use case?: Education - Learning & Teaching
Phone Number: +	

By Clicking "Sign Up", you agree to the [Privacy Policy](#)

Keep me informed with occasional updates about Databricks and Apache Spark™.

SIGN UP

3. Sélectionnez "Community Edition"

## COMMUNITY EDITION

For students and educational institutions

- Single cluster limited to 6GB and no worker nodes
- Basic notebooks without collaboration
- Limited to 3 max users
- Public environment to share your work

GET STARTED

By clicking "Get Started" for the Community Edition, you agree to the [Databricks Community Edition Terms of Service](#).

4. Contrôlez votre boîte mail et ouvrez le mail d'activation. Cliquez sur le lien d'activation et suivez les instructions.
5. Vérifiez l'accès à Databricks via l'URL suivant :  
<https://community.cloud.databricks.com/>

Enregistrez-vous avec votre adresse mail et le mot de passe fourni

6. Vérifiez que l'utilisateur est correctement activé et que le système fonctionne, comme suit :

