

# Mini PROJET : FasoDBpedia (exploration et extraction automatique de données issues du Portail National de Publication des Données Ouvertes)

## Contexte

Les administrations publiques et privées dans le cadre de leurs activités quotidiennes et de leurs services produisent et collectent une importante quantité de données. Cependant ces données sont difficilement accessibles et sous-exploitées par de tierces entités : personnes, organismes, applications. C'est ainsi que l'ouverture de ces masses de données permet de susciter la création de nouveaux secteurs à travers la création de nouveaux services et le développement d'applications pour le citoyen.

Les données ouvertes, sont des données auxquelles l'accès est public et libre de droit, tout comme leur exploitation. C'est ainsi qu'est née l'initiative des données ouvertes.

« L'initiative des données ouvertes ou open data est une philosophie publique et privée consistant à ouvrir les données d'intérêt public de manière à favoriser l'innovation ouverte dans le secteur public et privé, rendre des comptes aux citoyens et améliorer la transparence de l'action publique et enfin mettre à disposition du secteur privé des ressources permettant le développement de nouveaux services. L'open data est un mouvement composé d'acteurs militant pour l'ouverture des données. ». <sup>1</sup>

L'initiative **Open Data Burkina Faso** a mis en place une plateforme web dénommée **Portail National Publication des Données Ouvertes** (PNPDO). Le PNPDO est accessible à l'adresse suivante : <https://data.gov.bf/>. Ce portail est un référentiel de données où l'on a des producteurs (ou fournisseurs) qui fournissent des jeux de données sur divers thématiques (santé, éducation, infrastructures, mines, ...) du Burkina Faso.

C'est dans ce contexte que le projet **FasoDBpedia** vise à permettre l'exploration et l'extraction automatiques de données dérivées du Portail National de Publication de Données Ouvertes (PNPDO) en vue de leurs exploitations par des entités tierces en utilisant les technologies du Web sémantique et du Linked Open Data. Le but de la manœuvre est de proposer une version structurée et normalisée au format du web sémantique des contenus (producteurs, jeux de données, ressources, ...) du PNPDO. Nous souhaitons dans le cadre de ce projet FasoDBpedia (à l'instar de DBpedia pour les ressources de Wikipédia) mettre en œuvre l'exploration des données issues de la PNPDO. Cette exploration se fera au moyen de requête SPARQL. L'extraction et la combinaison de données issues de plusieurs sources (producteurs) en vue d'une exploitation.

## Objectif

L'objectif principal de ce projet **FasoDBpédia** est de réaliser des modèles (ontologie légère, schéma RDFS, description RDF) du PNPDO et de quelques jeux de données afin de permettre l'exploration et l'extraction automatiques de données issues du PNPDO. Vous devez donc proposer une version structurée et normalisée au format du web sémantique des contenus, ressources du PNPDO.

## Description du projet

- concevoir une ontologie (RDFS ou OWL) représentant les connaissances sur les métiers et familles d'emplois et leurs caractéristiques (échelle, catégorie). Possibilité de réutiliser les vocabulaires **foaf**, **skos**.
- sélectionner 6 jeux de données (dont on peut faire des liens, ou des requêtes liées en SPARQL

---

<sup>1</sup> <https://data.gov.bf/>

- pour chaque jeu de données, si le schéma RDFS n'est pas disponible utilisez le fichier (Excel ou csv) pour proposer un schéma RDFS de ce jeu de données ;
- décrire les ressources de ce jeu de données dans un format RDF en utilisant le schéma RDFS que vous avez utilisé/proposé
- Télécharger et utilisez le moteur Corese (à l'adresse suivante : <http://wimmics.inria.fr/doc/tutorial/corese.3.2.2.jar> vu en TP) pour mettre en pratique les objectifs du projet :
  - Requête SPARQL pour Lister :
    - l'ensemble des producteurs, les producteurs publics (administrations publiques), les producteurs privés ;
    - les thèmes ;
    - les jeux de données (nom, thèmes, producteurs,...)
  - A partir des jeux de données que vous avez choisis vous devez être capable d'élaborer des requêtes qui **combinent** des jeux de données issus de producteurs différents pour donner une information (*c'est la plus-value du Linked Open Data :=*)
- Rédiger un fichier WORD des résultats de vos requêtes
  - Mentionner l'objectif de chaque requête ;
  - Le code SPARQL de chaque requête
  - La capture d'écran de l'exécution dans CORESE.

## Travail à rendre :

- l'ontologie de la PNPDO, sérialisée en rdfs/xml ou en owl ;
- les schémas RDFS de chaque jeu de données (les 6 que vous auriez sélectionnées) ;
- les descriptions RDF de chaque jeu de données ;
- le rapport rédiger en WORD contenant :
  - les noms du binôme ;
  - la structuration du dossier projet, le nom des fichiers et leurs rôles ;
  - les requêtes , leurs objectifs, les réponses obtenues et l'illustration de l'exécution (screenshots) avec le moteur **CORESE**.

**NB : Le choix des jeux de données doit être bien réalisé de telle sorte, que ce ne soit pas tous des jeux de données qui proviennent du même producteur, mais aussi de sorte à ce qu'ils vous permettent de combiner ces données (par requête SPARQL) pour faire ressortir par exemple une information (du type : taux, ratio, rapport) réaliste sur la base des données issues de ces différentes sources(producteurs).**

**Projet à déposer sur votre compte GitHub !**

**Date limite de remise du projet : 6 février**

Envoyer le lien du dépôt GitHub de votre projet aux adresses suivantes avec pour objet « **Projet FasoDBpedia master SI-SAD** »:

- [sadouanouan.malo@gmail.com](mailto:sadouanouan.malo@gmail.com)
- [cedric.bere@gmail.com](mailto:cedric.bere@gmail.com)

Bonne chance !