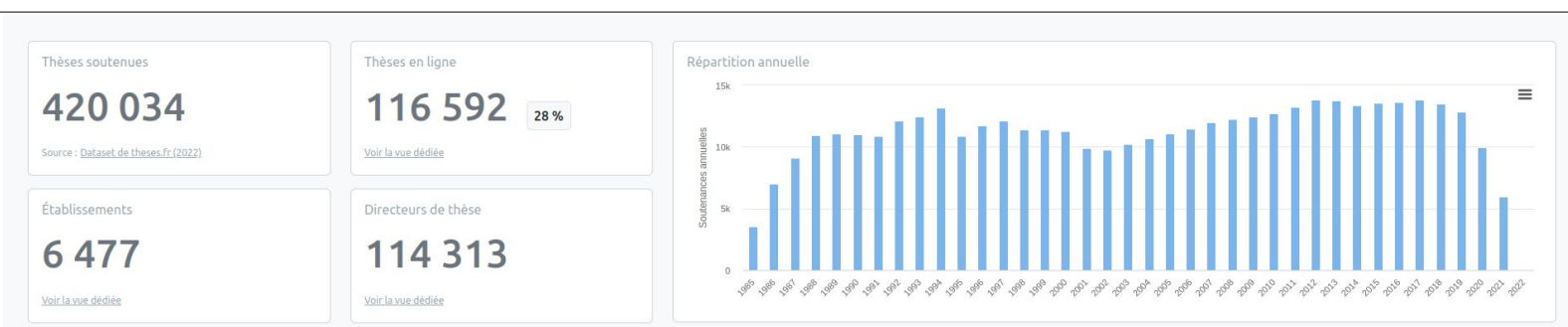


## CAHIER DES CHARGES

### PROJET « THESES VIZ »

Projet en lien avec les cours :  
R3.0.1 : Développement web [semestre 3]  
R4.Real.10 : Complément web [semestre 4]



Ce document descriptif du projet évoluera au cours de l'année.  
Il apportera tous les éléments attendus pour sa réalisation.

Rédacteur : Thomas FRESSIN  
Version : 16/09/22 - 10:58:37

## TABLE DES MATIÈRES

Description générale du projet.....	4
Organisme partenaire.....	4
Objectif général.....	4
Liens du projet avec le programme national.....	4
Savoirs de référence étudiés.....	4
Compétences ciblées.....	5
Apprentissages critiques ciblés.....	5
État de l’art.....	5
Phases générales du projet.....	6
Sécurité et protection des données.....	6
Rôles.....	6
Maîtrise d’ouvrage (MOA).....	6
Maîtrise d’oeuvre (MOE).....	7
Cibles de l’application.....	7
Spécifications générales.....	8
Fonctionnalités générales du site.....	8
Charte graphique.....	8
Charte éditoriale.....	8
Création et récupération de contenus.....	9
Description du jeu de données à importer.....	9
Contenus textuels.....	9
Contenus multimédias.....	9
Environnement global.....	10
Environnement de développement.....	10
Dépôt du nom de domaine et adresses email.....	10
Hébergement des données.....	10
Fonctionnalités.....	10
Moteur de recherche.....	10
Espace sécurisé de gestion d’alertes.....	11
Formulaire de contact.....	11
Sondage.....	11
Boutique en ligne.....	11
Quiz.....	11
Documentation.....	11
Réseaux sociaux.....	11
Flux RSS.....	11
Backoffice.....	11
Collecte d’information.....	12
Compatibilité navigateurs/écrans.....	12
Accessibilité.....	12
Logiciels utilisés pour le développement.....	12
Maquette du site.....	12
Référencement.....	12
Mesure d’audience.....	12
Mises à jour.....	12
Documentation et conformité.....	13

CNIL - RGPD.....	13
Formation.....	13
Dépôt git.....	13
Livraison.....	14
Planning de livraisons.....	14
Livrables attendus à chaque phase.....	14
Notation.....	15
Budget.....	16
Néant pour ce projet.....	16
Annexes.....	17
Annexe 1 – Champs minimaux à importer.....	17
Annexe 2 – Tableau de bord.....	18

# DESCRIPTION GÉNÉRALE DU PROJET

## Organisme partenaire

L'Abes (Agence bibliographique de l'enseignement supérieur) est un établissement public à caractère administratif (EPA) sous tutelle du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI).

Créée en 1994 par décret interministériel pour mettre en œuvre le Sudoc (Système Universitaire de Documentation), l'Abes joue un rôle important, par ses actions de mutualisation, auprès des bibliothèques universitaires et de recherche françaises. Son siège est situé à Montpellier.

L'agence met en œuvre le catalogue des thèses via son application « [theses.fr](https://theses.fr) ».

## Objectif général

Sur la base d'une extraction (fournie) ou de la totalité de la base de données et métadonnées des thèses soutenues inscrites au catalogue des thèses « [theses.fr](https://theses.fr) » (téléchargeable au format JSON téléchargeable sur [data.gouv.fr](https://data.gouv.fr)), réaliser une application web (langage de programmation PHP objet obligatoire) présentant un tableau de bord permettant une nouvelle expérience de visualisation des données que la version actuelle du catalogue des thèses « [theses.fr](https://theses.fr) ».

Ce projet se déroulera sur les deux semestres de 2<sup>e</sup> année de BUT INFORMATIQUE.

## Liens du projet avec le programme national

### Savoirs de référence étudiés

- Programmation web PHP (authentification, service web, opendata, visualisation)
- Sensibilisation à la sécurité web
- Sensibilisation à la sécurité des applications
- Protection des données personnelles
- Mise en œuvre de l'ergonomie
- Mise en œuvre de l'accessibilité

## **Compétences ciblées**

- Développer une solution informatique pour un client
- Proposer des applications informatiques optimisées
- Installer, configurer, mettre à disposition, maintenir en conditions opérationnelles des infrastructures, des services et des réseaux et optimiser le SI d'une organisation
- Concevoir, générer, administrer et exploiter les données de l'entreprise et mettre à disposition toutes les informations pour un bon pilotage de l'entreprise.
- Satisfaire les besoins des utilisateurs au regard de la chaîne de valeur du client, organiser et piloter un projet informatique avec des méthodes classiques ou agiles.

## **Apprentissages critiques ciblés**

- Élaborer et implémenter les spécifications fonctionnelles et non fonctionnelles à partir des exigences
- Appliquer des principes d'accessibilité et d'ergonomie
- Adopter de bonnes pratiques de conception et de programmation
- Comprendre les enjeux et moyens de sécurisation des données et du code
- Sécuriser les services et données d'un système
- Organiser la restitution de données à travers la programmation et la visualisation
- Manipuler des données hétérogènes
- Évaluer l'impact ~~environnemental~~ et sociétal des solutions proposées
- Assurer la sécurité des données
- Formaliser les besoins du client et de l'utilisateur

## **État de l'art**

A ce jour, le site theses.fr existe depuis plus de dix années et n'intègre plus de nouvelles fonctionnalités.

A ce jour, aucun site ne semble réexploiter les données publiées en ligne de façon ouverte du catalogue des thèses.

L'intérêt du projet est donc de proposer une nouvelle expérience de navigation dans ces données.

## Phases générales du projet

1. Concevoir puis développer les classes utiles et la base de données associée
2. Concevoir puis développer un script d'import des données du catalogue de thèses vers la base de données utilisée dans l'application
3. Concevoir puis développer les fonctionnalités de recherche de thèses
4. Concevoir puis développer la visualisation des données sous la forme d'un tableau de bord dynamique, c'est-à-dire en lien avec un formulaire de recherche (en utilisant les fonctionnalités du point ci-dessus)
5. Concevoir puis développer une interface permettant de gérer des alertes sur mots-clés
6. Concevoir puis développer des interactions avec d'autres bases de données ouvertes

Durant toute la durée du projet, l'étudiant veillera :

- à la sécurité et à la protection des données « *by design* » (voir la section « sécurité et protection des données ») ;
- à ce que le rendu soit *responsive* ;
- à la mise en application d'une ergonomie.

## Sécurité et protection des données

L'application développée devra respectée les recommandations techniques en matière de sécurité et de protection des données précisées dans le [\*Guide de l'équipe de développement de la CNIL\*](#) (cnil) et présentées durant le cours.

## Rôles

### Maîtrise d'ouvrage (MOA)

Rôle global de la maîtrise d'ouvrage :

- valider les phases de choix (ergonomie, contenus textuels),
- valider le respect du cahier des charges,
- veiller au respect des délais,
- fournir les contenus de base (textes, plaquettes, logos, images, photos).

Interlocuteurs :

- Thomas FRESSIN, enseignant
- L'ABES (en fin de projet)

## **Maîtrise d'oeuvre (MOE)**

Rôle théorique de la maîtrise d'oeuvre :

- concevoir et réaliser l'application,
- - respecter les délais,
- proposer un planning de réalisation en accord avec le client\*

\*Ce point sera cadencé par l'enseignant.

Interlocuteurs :

- Chaque étudiant

## **Cibles de l'application**

<b>Cible</b>	<b>Caractéristiques</b>	<b>Importance</b>	<b>Centres d'intérêts</b>
Chercheurs (doctorants, docteurs) et futurs chercheurs	Haut niveau d'instruction, 20-90 ans, bonnes connaissances des outils de recherche	Primaire	Information sur les thèses, le réseau de chercheurs, téléchargement des thèses
Grand public, journalistes	20-90 ans	Secondaire	Information

# SPÉCIFICATIONS GÉNÉRALES

## Fonctionnalités générales du site

Le projet doit permettre :

- la recherche de thèses soutenues *a minima* par titre et par nom/prénom
- la visualisation des données sur les thèses, associée à la recherche en cours
- la visualisation doit être « *responsive* » sur un appareil type PC, smartphone ou tablette

## Charte graphique

L'étudiant est invité à utiliser [Bootstrap](#) pour gagner en productivité.

L'étudiant peut largement s'inspirer de thèmes existants de tableaux de bord ([exemple](#)).

Le développement de la charte graphique et de l'ergonomie de l'application est à la charge de l'étudiant.

L'application devra comprendre *a minima* :

- Un header, un cadre « Moteur de recherche »
- Un cadre « Espace de visualisation des données sous format graphique », les données étant filtrées selon la requête
- Un cadre « Liste des thèses » filtrées selon la requête
- En footer, tout élément supplémentaire intéressant la description de l'application [à la discrétion de l'étudiant]

## Charte éditoriale

Néant.



## Création et récupération de contenus

### Description du jeu de données à importer

Lien vers le jeu de données :

<https://www.data.gouv.fr/fr/datasets/theses-soutenues-en-france-depuis-1985/>

Le jeu de données utilisé contient les métadonnées des thèses de doctorat françaises soutenues depuis 1985.

Les données sont issues de [theses.fr](https://theses.fr), le moteur de recherche des thèses de doctorat françaises. Il recense l'ensemble des thèses de doctorat soutenues en France depuis 1985. Les données sont produites par les établissements habilités à délivrer le doctorat et sont placées sous leur responsabilité.

Seules les thèses de doctorat françaises soutenues depuis 1985, telles que définies par [l'arrêté du 25 septembre 1985](#), [l'arrêté du 7 août 2006](#) et [l'arrêté du 25 mai 2016](#) (mis à jour par [l'arrêté du 26 août 2022](#)), font l'objet d'un signalement sur [theses.fr](https://theses.fr) (et sont présentes dans ce jeu de données). Ce signalement constitue une obligation réglementaire.

Ce jeu de données est disponible en trois formats : [CSV](#), [json](#) et [ndjson](#). Il est accompagné d'un [guide de réutilisation](#).

Les données au format JSON doivent être importées via un script PHP intitulé « script\_import.php », enregistré dans un dossier « scripts/ » situé à la racine du projet.

Voir annexe 1 pour les champs concernés par l'import.

### Contenus textuels

Aucun contenu textuel fourni dans le cadre de ce projet.

### Contenus multimédias

Aucun contenu multimédia fourni dans le cadre de ce projet.

En cas d'utilisation de contenus multimédias, l'étudiant doit s'assurer des droits d'utilisation de ces contenus (libres).

## Environnement global

### Environnement de développement

- Environnement de développement adapté au projet (serveur web, base de données, GSBDD)
- HTML, CSS
- PHP objet v.7 mini, sans utiliser de framework
- Base de données MySQL ou SQLite
- [Bootstrap](#)
- Opendata (theses.fr)
- Visualisation via la librairie [highcharts](#)

### Dépôt du nom de domaine et adresses email

Néant pour ce projet.

### Hébergement des données

Dans le cadre du projet, l'hébergement se réalisera sur le serveur web de l'université ou bien tout autre serveur web.

Dév.... prod

Toutefois, les étudiants livreront leur site sur un serveur web (de l'université ou autre) en veillant à y intégrer une authentification htaccess, avec pour identifiants :

- login : [thomas.fressin@univ-eiffel.fr](mailto:thomas.fressin@univ-eiffel.fr)
- password : [gTF9#9#xsgr9D@rWxzat1xE](#)

Le projet devra être supprimé avant le mois de juillet 2023.

## Fonctionnalités

### Moteur de recherche

Le moteur de recherche, placé en header du site, doit permettre de recherche des thèses *a minima* à partir des champs « titre » (en langue FR) et « nom/prénom » des auteurs des thèses.

L'étudiant est libre d'implémenter d'autres champs s'il le souhaite.

## **Espace sécurisé de gestion d'alertes**

L'étudiant doit créer un espace sécurisé permettant aux utilisateurs de s'enregistrer et de gérer des alertes sur mots-clés.

Un script doit pouvoir envoyer des alertes en cas d'intégration de nouveau contenu.

Dans la pratique, dans le cadre du projet, l'enseignant doit pouvoir se connecter avec ces identifiants :

- login : [thomas.fressin@univ-eiffel.fr](mailto:thomas.fressin@univ-eiffel.fr)
- password : gTF9#9#xsgr9D@rWxzat1xE

Puis doit pouvoir ajouter/modifier/supprimer ses alertes sur mots-clés.

L'interface doit proposer un bouton « envoyer l'alerte », qui envoie un courriel incluant un tableau listant les thèses liées à la requête de l'alerte.

## **Formulaire de contact**

L'étudiant est libre d'en créer un ou non.

## **Sondage**

Néant pour ce projet.

## **Boutique en ligne**

Néant pour ce projet.

## **Quiz**

Néant pour ce projet.

## **Documentation**

A la discrétion de l'étudiant.

## **Réseaux sociaux**

Néant.

## **Flux RSS**

Néant.

## **Backoffice**

Néant.

## **Collecte d'information**

L'étudiant est libre de créer des tables de traces (consultation, exceptions, ...).

## **Compatibilité navigateurs/écrans**

Le site doit être *responsive* et lisible sur tout navigateur ou écran.

## **Accessibilité**

Conformément à la [réglementation](#), le site respectera le Référentiel Général d'Accessibilité pour les Administrations.

Ainsi, un lien « Accessibilité » sera présent en footer sur le site pour permettre aux utilisateurs mal voyants d'agrandir la police afin de faciliter la lecture

## **Logiciels utilisés pour le développement**

Libre.

## **Maquette du site**

L'étudiant peut largement s'inspirer de thèmes existants de tableaux de bord ([exemple](#)).

## **Référencement**

Aucun référencement n'est ici à prévoir.

Un fichier robots.txt, bien configuré, doit être placé à la racine du projet, pour empêcher l'indexation du site par les robots d'indexation.

## **Mesure d'audience**

En cas d'utilisation d'outil de mesure d'audience, soit en réaliser un from scratch, soit utiliser une solution autorisée par la CNIL (voir la page [Cookies : solutions pour les outils de mesure d'audience](#) qui liste les recommandations et les solutions conformes à la protection des données selon la CNIL).

## **Mises à jour**

L'étudiant peut concevoir et développer un script d'import de données qui permettra d'intégrer de nouvelles thèses et mettre à jour les thèses déjà connues.

## **Documentation et conformité**

### **CNIL - RGPD**

En cas d'utilisation de données personnelles ou de cookies, l'étudiant doit veiller à la conformité de l'application au RGPD.

Il n'est pas demandé à l'étudiant de rédiger un registre de traitement ou une AIPD. Mais cela devrait être réalisé dans un vrai projet.

### **Formation**

L'outil devra être suffisamment ergonomique pour qu'aucune formation des utilisateurs ne soit nécessaire.

### **Dépôt git**

Le dépôt git utilisé pour déposer le code doit avoir une visibilité privée.

# LIVRAISON

## Planning de livraisons

Le projet sera livré par l'étudiant à l'enseignant en plusieurs étapes intermédiaires, conformément aux phases générales du projet. Il serviront de points d'avancée et de notation.

Des livraisons sont à prévoir lorsque :

1. Les classes des objets PHP et l'ossature principale de la base de données sont développées
2. Le script d'import des données du catalogue de thèses vers la base de données est opérationnel sur l'échantillon de données fournies
3. Les fonctionnalités de recherche de thèses sont opérationnelles
4. La visualisation des données sous la forme d'un tableau de bord dynamique (en lien avec les fonctionnalités de recherche) est opérationnelle
5. L'interface permettant de gérer des alertes sur mots-clés
6. Des interactions avec d'autres bases de données ouvertes sont opérationnelles
7. Livraison finale à la date d'échéance, quelque soit l'avancée du projet

## Livrables attendus à chaque phase

L'étudiant fournira à chaque avancée d'une phase de son projet, par l'envoi d'un courriel à [thomas.fressin@univ-eiffel.fr](mailto:thomas.fressin@univ-eiffel.fr) :

- l'url du site accessible ;
- l'url du dépôt git utilisé (veiller à transmettre les droits de lecture à l'enseignant au courant du mois d'octobre), qui devra contenir un dump de la base de données dans un dossier « bdd/ » à la racine du projet ;
- le cas échéant, tout document supplémentaire lié au projet et que l'étudiant souhaite que l'enseignant examine.

La date de livraison finale est fixée ainsi :

- pour les étudiants du groupe A : 15 novembre 2022 à 23h59
- pour les étudiants du groupe A : 30 novembre 2022 à 23h59

## Notation

L'enseignant intègre dans sa notation :

- la qualité du code, les commentaires associés, le respect de la sécurité informatique, ...
- la forme, l'ergonomie et l'accessibilité du rendu final
- la montée individuelle en compétences et l'investissement personnel
- le respect des délais
- la qualité des présentations aux revues de code

Tout retard de l'étudiant sur la livraison finale sera sanctionné (-2 points par demi-journée de retard).

Les trois meilleurs projets seront présentés à la cheffe de projet du site theses.fr de l'ABES pour être départagés. Il est proposé au lauréat de publier le site (modulo de potentielles adaptations/modifications supplémentaires), sous la direction de l'enseignant, comme « réutilisation » officielle des données sur le site data.gouv.fr Un bonus de +1 points sera attribué au lauréat qui aura publié ce site ainsi.

## BUDGET

Néant pour ce projet.

Toutefois l'étudiant doit créer un fichier de reporting du temps total passé pour chaque grand module de ce projet.

Ce fichier doit être présent à la racine du site et accessible via un lien intitulé « reporting », placé dans le footer du site.



# ANNEXES

## Annexe 1 – Champs minimaux à importer

Champs minimaux à créer pour le projet (dans le code php objet + base de données) :

Pour les thèses :

- id
- en\_ligne
- auteurs
- date\_soutenance
- directeurs\_these
- discipline
- embargo
- etablissements\_soutenance
- langue
- membres\_jury
- nnt
- oai\_set\_specs
- president\_jury
- rapporteurs
- resume
- soutenue
- sujets
- these\_sur\_travaux
- titre

Pour les personnes (docteurs, directeurs de thèse, membres du jury) :

- id
- nom
- prénom
- idref

Jeu de test à importer à importer (1361 thèses) : [https://tfressin.fr/thesesviz/extract\\_theses.json](https://tfressin.fr/thesesviz/extract_theses.json)

## Annexe 2 – Tableau de bord

### Chiffres minimum à faire ressortir :

- Nombre de thèses soutenues
- Nombre de thèses en ligne
- Nombre d'établissements
- Nombre de directeurs de thèse

### Graphiques minimum à intégrer :

- thèses en ligne/pas en ligne (camembert)
- évolution du nombre de thèses en ligne par année
- évolution par année du nombre de thèses (pour voir l'évolution du nombre de thèses soutenues chaque année depuis 1985)
- nombre de thèses soutenues par mois (pour voir les mois où il y a le plus de soutenances)
- par discipline (pour voir les disciplines majoritaires)
- par langue
- thèses sous embargo
- nuage de tags des sujets

### Visualisation avancée :

- Cartographie des établissements de soutenance
- Graphique du réseau d'un doctorant

Toute la visualisation doit être réalisée avec la librairie Highcharts.

### Liste de thèses :

Le tableau de bord doit présenter un aperçu d'une liste de thèses (avec un top 10 ou 20) basée sur les filtres en cours. Il doit être possible de cliquer sur la thèse est de :

- arriver sur une fiche descriptive de la thèse (soit via theses.fr, soit sur votre appli)
- si la thèse est en ligne, la télécharger pour y accéder directement