

Ting WANG

DUT STID 2^{ème} année

Années universitaires 2019 – 2021

Maître d'apprentissage : Madame Séverine MAYO-SIMBSLER

Tuteur académique : Monsieur Guillaume SANTINI

RAPPORT FINAL D'ALTERNANCE

MISSIONS AU SEIN DU DÉPARTEMENT DES ÉTUDES STATISTIQUES DE LA RECHERCHE SOUS-DIRECTION DES SYSTÈMES D'INFORMATION ET DES ÉTUDES STATISTIQUES (SIES)



**MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION**

*Liberté
Égalité
Fraternité*



Remerciements

C'est une réelle opportunité pour moi de pouvoir effectuer ma deuxième année de DUT STID en alternance en rejoignant le Département des études statistiques de la recherche.

Avant d'entrer dans le cœur du sujet, je souhaite remercier toute l'équipe du département recherche pour la qualité de leur accueil, bien que nous sommes dans un contexte particulier de crise sanitaire, je n'ai pas eu de mal à m'intégrer.

Je souhaite également remercier toute l'équipe pédagogique du département STID pour la qualité de leurs enseignements durant mes deux années de DUT.

Je tiens à remercier particulièrement Madame Séverine MAYO-SIMBSLER, mon maître d'apprentissage pour sa bienveillance. Elle a veillé à ma bonne intégration et au bon déroulement de cette année d'apprentissage au sein du département.

Je tiens également à remercier Monsieur Thomas BALCONE, qui a veillé au bon déroulement des missions et qui m'a été un soutien fort lorsqu'un blocage survient durant un projet.

Enfin, je remercie Monsieur Guillaume SANTINI, mon tuteur académique de m'avoir suivi durant cette année d'alternance, et qui a donné de précieux conseils pour la rédaction des rapports.

Glossaire

CHRU : Centres hospitaliers régionaux et universitaires

CHU : Centres hospitaliers universitaires

CISAD : Concepteur développeur d'applications pour la statistique

CLCC : Centre de lutte contre le cancer

DERDE : Dépenses extérieures de recherche et développement des entreprises

DIRD : dépenses intérieures de recherche et développement

DMCSI : La division des Méthodes Appliquées de l'Econométrie

DNRD : dépenses nationales de recherche et développement

EESRI : Etat de l'enseignement supérieur, de la recherche et de l'innovation en France

EPA : Etablissement public à caractère administratif

EPIC : Etablissement public à caractère industriel et commercial

ETP : Equivalent temps plein

INSEE : Institut national de la statistique et des études économiques

ISBL : Institution sans but lucratif

GIP : Groupement d'intérêt général

Manuel de Frascati : référence méthodologique en matière de recueil et de statistique de R&D

MESRI : Ministère de l'enseignement supérieur, de la recherche et de l'innovation

NAF : Nomenclature d'activités française

NF : Notes flash

NI : Notes d'information

OCDE : Organisation de coopération et de développement économique

OpenData : Plateforme ouvertes des données sur l'enseignement supérieur, la recherche et innovation

PP : Personne physique

R&D : Recherche et développement

SSM : Service statistique ministériel

SOMMAIRE

Remerciements	1
Glossaire	2
1 Introduction	5
2 Présentation de l'entreprise	7
2.1 Organisation du Ministère	7
2.2 Organisation de la Sous-direction SIES	10
2.3 Le département recherche	11
2.4 Organisation du département	15
2.4.1 L'équipe entreprise	16
2.4.2 L'équipe administration	17
3 Les missions	18
3.1 La brochure EESRI 2021	18
3.1.1 La fiche 35 : Les moyens humains de la R&D	20
3.1.2 La fiche 36 : La parité dans la recherche	22
3.2 Intégration des données du département vers l'OpenData	26
3.3 Les réponses à la demande	32
3.4 Le Funathon	34
3.4.1 Présentation des données	35
3.4.2 Analyse	38
4 Le bilan	51
Conclusion	54
Annexes	56

INTRODUCTION

Le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) a été créé en 1974 sous le gouvernement Chirac. Il a pour but de mettre en pratique les politiques du gouvernement relative aux accès à l'enseignement supérieur, à l'acquisition des savoirs et aux développements de la recherche et de l'innovation scientifique et technologique en France.

Le MESRI intervient dans différents domaines de l'enseignement et de la recherche. Il est compétent dans la définition de la politique de l'espace. Il est également acteur des réformes majeures des projets pédagogiques des établissements d'enseignement supérieur.

Il est aussi l'un des plus gros bénéficiaires du budget de l'Etat. Le budget accordé au MESRI en 2020 est de 25,49 milliards d'euros (près de 6% du budget de l'Etat voté en 2020), ce qui le place en 4^e position. Il a donc pour rôle d'attribuer les ressources et les moyens dans le cadre de la mission interministérielle Recherche et enseignement supérieur en collaboration avec d'autres ministères.

Il doit définir et mettre en place les projets d'investissement d'avenir, tels que dans les domaines du numérique, dans la recherche fondamentale, dans le développement durable etc.

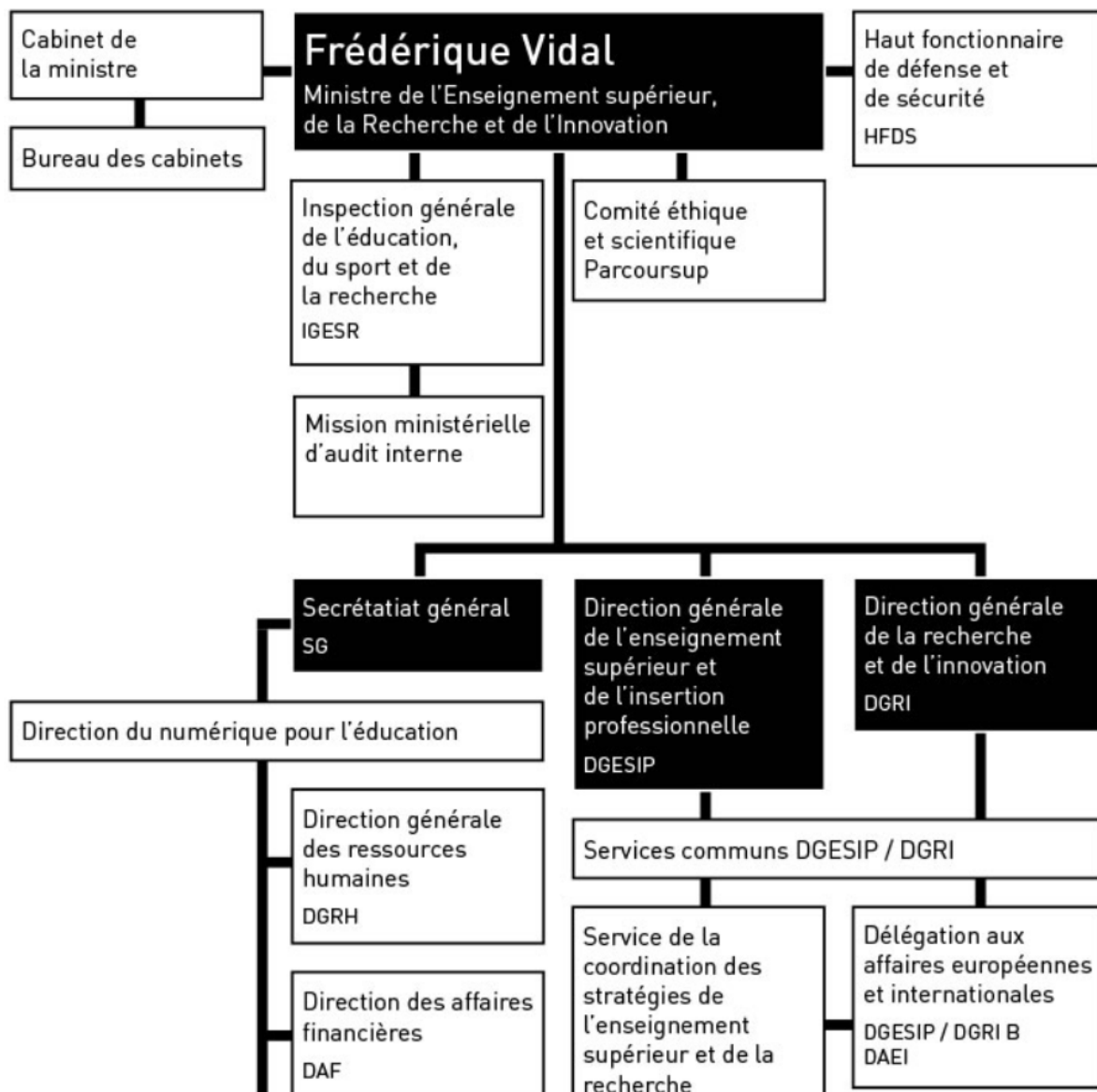
Il est également responsable de l'innovation dans divers domaines en collaboration avec les organismes de recherche sous sa tutelle.

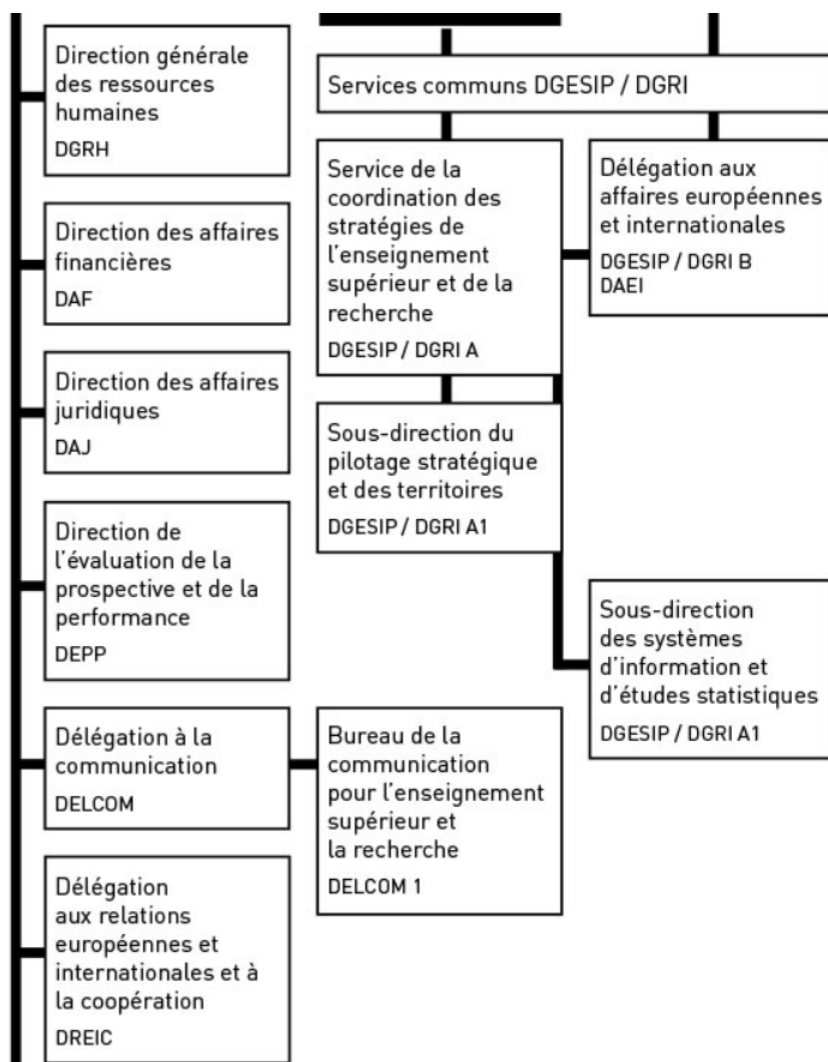
Il est chargé de promouvoir l'enseignement et la recherche française à l'international, en fournissant des données statistiques aux organismes compétents.

Il diffuse la culture scientifique, technologique et industrielle auprès du grand public à travers divers événements scientifiques organisés tout au long de l'année.

2 Présentation de l'entreprise

2.1 Organisation du Ministère





Organigramme de l'Administration Centrale du MESRI

Je vais maintenant présenter l'organisation du MESRI. Dans cet organigramme nous avons tout en haut la ministre actuelle Frédérique VIDAL. Le poste de ministre du MESRI n'est pas toujours créé. Selon le gouvernement, ce poste peut être occupé conjointement par le ministre de l'Education nationale. Par exemple, sous le gouvernement Valls et celui de Cazneuve, entre 2014 et 2017, ce poste a été occupé par l'ancienne ministre de l'Education nationale Najat VALLAUD-BELKACEM.

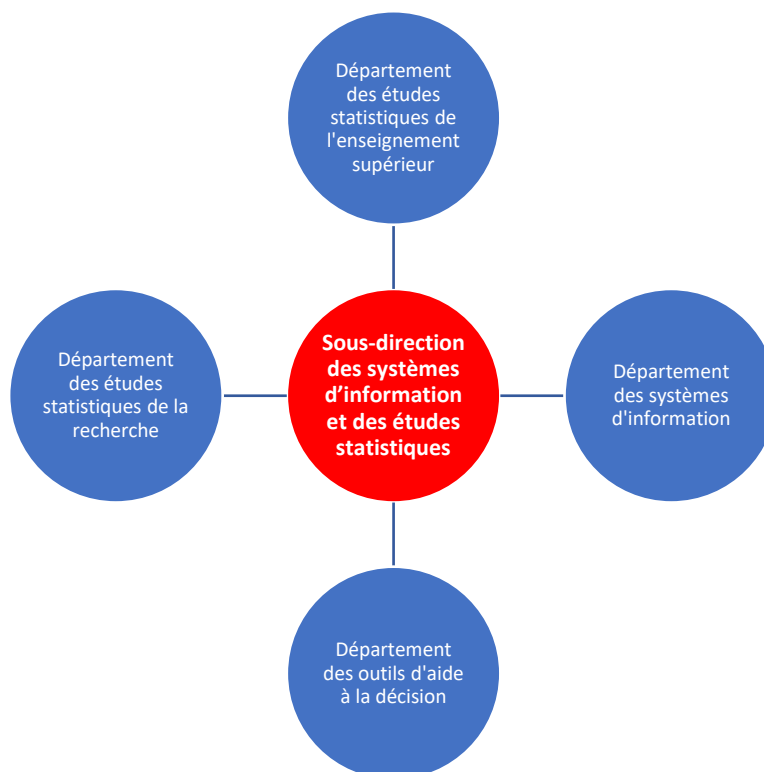
Le MESRI s'appuie sur trois inspections générales pour mener à bien ses missions. L'inspection générale de l'éducation, du sport et de la recherche (IGESR) est

directement sous les ordres du MESRI. Il partage avec le ministère de l'Education nationale l'inspection générale des bibliothèques (IGB) et l'inspection générale de l'éducation nationale (IGEN).

Il y a également plusieurs directions générales et un secrétariat général (SG) qu'il a directement l'autorité, dont La direction générale de l'enseignement supérieur et de l'insertion professionnel (DGESIP) et la direction générale de la recherche et de l'innovation (DGRI).

Les directions générales partagent plusieurs services et sous-directions en communs. Le service de la coordination des stratégies de l'enseignement supérieur et de la recherche, la délégation des affaires européennes et internationales (DAEI), la sous-direction du pilotage stratégique et des territoires, ainsi que la sous-direction dans laquelle j'effectue mon apprentissage, la sous-direction des systèmes d'informations et des études statistiques (SIES).

2.2 Organisation de la Sous-direction SIES



Les différents départements au sein de la Sous-direction SIES

La Sous-direction des systèmes d'information et des études statistiques (SIES) est un service statistique ministériel (SSM) de l'enseignement supérieur et de la recherche créé en 2009. L'actuelle sous-directrice est Isabelle KABLA-LANGLOIS. Ce SSM est constitué de quatre départements : le département des études statistiques de l'enseignement supérieur, le département des études statistiques de la recherche, le département des systèmes d'information et le département des outils d'aide à la décision.

La sous-direction a pour mission de produire des données statistiques, et réalise des études statistiques, afin d'évaluer les politiques du gouvernement sur l'enseignement supérieur telles que l'orientation, le parcours et la réussite des étudiants. Elle produit également des études statistiques sur la R&D (Recherche

et développement) dans les entreprises ainsi que dans les organismes de recherche et dans les universités.

Ces données statistiques sont collectées généralement une fois par an, directement auprès des institutions concernées, qu'elles soient publiques ou privées.

A partir de ces études quantitatives, elle élabore des outils de mesure de la performance et d'aide à la décision au sein du ministère sous forme de tableaux de bord, de datavisualisations et d'informations géolocalisées.

Elle transmet des données statistiques réglementaires aux institutions européennes et internationales (Eurostat, OCDE, ...) dans le domaine de l'enseignement supérieur et de la recherche.

2.3 Le département recherche

Le département des études statistiques de la recherche est l'un des quatre départements du SIES. Il a été créé en Août 2016 en tant que département autonome.

La mission principale du département est d'assurer la production des agrégats économiques tels que la DIRD (Dépenses intérieures de R&D) et la DNRD (Dépenses nationales de R&D).

La R&D est définie par le manuel de Frascati, un outil de référence pour les statisticiens du monde de la recherche et innovation établi par l'OCDE (Organisation de coopération et de développement économique). Les données recueillies se focalisent principalement dans les trois grands domaines : la

recherche fondamentale, la recherche appliquée et le développement expérimental.

Pour faire ces études, le département réalise plusieurs enquêtes statistiques auprès des entreprises et des administrations. Plus de 11 000 entreprises sont interrogées chaque année. Les données des organismes de recherche, dont 40 EPIC (Etablissement public à caractère industriel et commercial), EPA (Etablissement public à caractère administratif), les établissements d'enseignement supérieur et de recherche publics et privés, les CHU (Centres hospitaliers universitaires), CHRU (Centres hospitaliers régionaux et universitaires) et CLCC (Centre de lutte contre le cancer) ainsi que les associations et fondations de recherche sont également interrogés annuellement.

Toutes les entités qui investissent dans la R&D, quelles que soient leur taille et leur localisation géographique (métropole et Outre-mer) sont concernées par ces enquêtes annuelles. Elles doivent fournir des données concernant leurs dépenses, leurs ressources, leurs effectifs et leurs clientèles.

Le département produit les statistiques nationales sur la R&D en France et les envoie à l'OCDE et à l'Eurostat, afin d'informer l'Union européenne sur les moyens (humains et financiers) consacrés à la R&D. Ces statistiques permettent de faire un état des lieux sur les dépenses intérieures et extérieures. Elles permettent également de savoir les effectifs de chercheurs et de personnels de soutien à la R&D ainsi que les financements reçus pour ces moyens.

Le département est aussi producteur de publications. Il produit et contribue à plus de 50 publications dont les NI (Notes d'information) et NF (Notes flash) (cf. [annexe 1 et 2](#)).

Le département est auteur de deux publications à large diffusion : « *EESRI* » et la brochure « *Vers l'égalité femmes-hommes ? Chiffres clés* » qui est diffusée tous les ans, le 8 Mars à l'occasion de la journée internationale des droits des femmes.

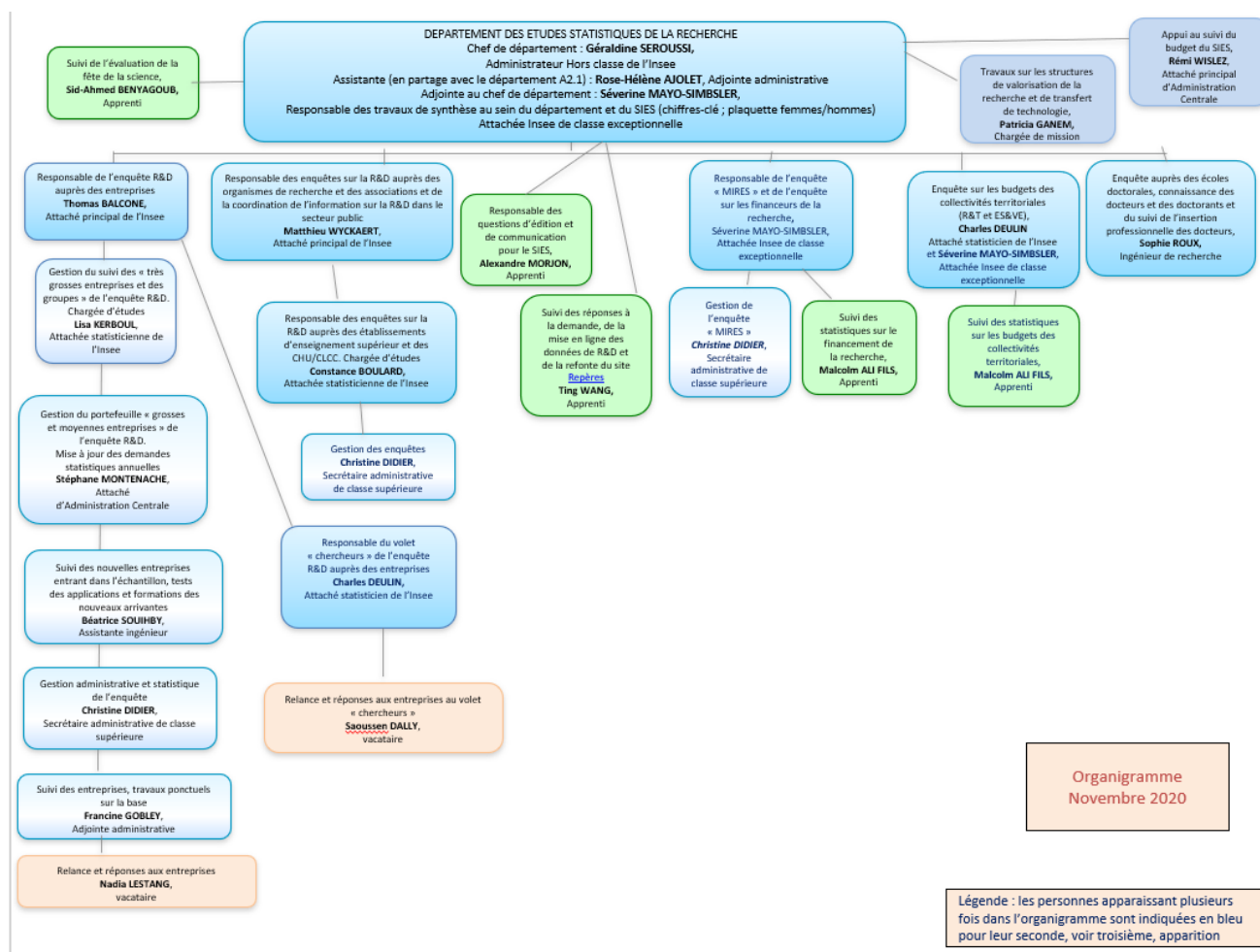


*La brochure l'Etat de l'enseignement supérieur de la recherche et de l'innovation en France
(version 2019)*



La brochure Vers l'égalité femmes-hommes ? Chiffres clés (version 2021)

2.4 Organisation du département



Organigramme du département de la Recherche A2.2

Voici un organigramme du département. Le département de la recherche est composé de 20 personnes, dont 4 apprentis et 2 vacataires. Il y a une équipe sur les statistiques de R&D côté entreprise et une autre sur les statistiques de R&D côté administration.

La secrétaire, Madame Rose-Hélène AJOLET travaille également pour le département de l'enseignement supérieur.

Madame Géraldine SEROUSSI est l'actuelle cheffe du département recherche. Elle est également administrateur hors classe de l'Insee. Elle a pour mission de :

- La production et la diffusion des agrégats nationaux sur la R&D
- Suivre les dispositifs sur l'insertion professionnelle des docteurs

Madame Séverine MAYO-SIMBSLER est l'adjointe à la cheffe du département. Elle est aussi ma maîtresse d'apprentissage. Elle est également attachée hors classe de l'Insee. Elle a pour mission de :

- Manager l'ensemble du département
- Assurer l'intérim du responsable du département en cas d'absence de celle-ci
- La coordination des travaux de synthèse du côté recherche
- La mise à jour de la connaissance du département sur la thématique recherche
- Le suivi des questions budgétaires et financières de la R&D
- Assurer le bon déroulement de l'enquête MIRES (Mission interministérielle recherche et enseignement supérieur) et l'enquête sur les financeurs de la recherche.

2.4.1 L'équipe entreprise

Monsieur Thomas BALCONE est responsable de l'enquête R&D auprès des entreprises. Il est également attaché principal de l'Insee. Il a pour mission de :

- La mise en œuvre de l'enquête réalisée auprès des entreprises
- La mise en place des travaux méthodologiques (constitution du questionnaire, échantillonnages, redressements, pondérations) en rapport avec l'enquête
- La collecte des données basant sur les outils informatiques
- Le suivi de l'évolution des enquêtes

L'équipe entreprise est composée de 9 personnes, Monsieur Balcone est soutenu par son équipe pour les différentes missions. Dans l'équipe on compte 37.5% d'attaché de l'Insee soit trois personnes, un attaché de l'administration centrale, une assistante ingénieure, deux personnels administratifs et deux vacataires.

Ils ont pour missions de :

- La gestion, de la réalisation et du suivi des enquêtes sur la R&D auprès des entreprises et des groupes
- De la mise à jour des demandes statistiques annuelles
- De tester les applications et de former les nouveaux arrivants
- De la gestion administrative et statistique de l'enquête
- De la gestion de l'enquête MIREs
- D'effectuer des travaux ponctuels sur la base de données
- De réaliser l'enquête sur les budgets des collectivités territoriales

2.4.2 L'équipe administration

Monsieur Matthieu WYCKAERT est responsable des enquêtes sur le secteur public. Il est également attaché principal de l'Insee. Il a pour mission :

- D'assurer les enquêtes de R&D auprès des organismes publics et des établissements d'enseignement supérieur et de recherche
- De réaliser la consolidation des résultats des enquêtes portant sur le secteur public
- De la coordination des travaux réalisés par les autres responsables d'enquêtes.
- La mise en cohérence des flux de financements croisés entre le secteur public et privé

L'équipe administration est composée de 3 personnes, Monsieur WYCKAERT est soutenu par son équipe pour les différentes missions. Dans l'équipe on compte deux attachés de l'Insee et une gestionnaire des enquêtes.

Ils ont pour missions de réaliser les enquêtes auprès des établissements d'enseignement supérieur et des CHU, CLCC.

Nous avons également des apprentis qui travaillent directement avec la cheffe du département ou l'adjointe à la cheffe, notamment sur les questions du suivi et de l'évaluation des enquêtes liées à la fête de la science ainsi que toutes les questions relatives à l'édition des publications et à la communication du département.

3 Les missions

3.1 La brochure EESRI 2021



La brochure EESRI (version 2021)

La brochure « l'Etat de l'enseignement supérieur, de la recherche et de l'innovation » (EESRI) est une publication annuelle. Elle sera à sa 14^e édition pour la version 2021. Elle présente la situation de l'enseignement supérieur et de la recherche et innovation en France à travers des données chiffrées et des illustrations graphiques. Une vision sur l'international est également abordée. Elle compare la France aux autres pays de l'OCDE sur l'enseignement supérieur et sur la recherche. Elle comporte une préface Ministre et un résumé qui résume les principales évolutions et changements par rapport aux années précédentes.

Les 53 thématiques de l'édition 2021

Sommaire de la 14^e édition de l'Etat de l'Enseignement supérieur, de la Recherche et de l'Innovation

- [préface](#)
- [l'état de l'Enseignement supérieur, de la recherche et de l'innovation - résumé](#)

Enseignement supérieur

- 01 - [la dépense d'éducation pour l'enseignement supérieur](#)
- 02 - [la dépense pour l'enseignement supérieur dans les pays de l'OCDE](#)
- 03 - [l'aide sociale aux étudiants](#)
- 04 - [les personnels enseignants de l'enseignement supérieur public sous tutelle du MESRI](#)
- 05 - [les personnels non-enseignants de l'enseignement supérieur public sous tutelle du MESRI](#)
- 06 - [les salaires des personnels des universités et des EPST](#)
- 07 - [qualification et recrutement des enseignants-chercheurs](#)
- 08 - [les nouveaux bacheliers et leur entrée dans les filières de l'enseignement](#)

Une partie du sommaire de la brochure (version 2021)

L'ouvrage comporte une centaine de pages avec plus d'une cinquantaine de thèmes abordés. Chaque thème est sur une double page, avec la synthèse à gauche et les illustrations à droite. Ces thèmes sont divisés selon qu'ils concernent l'enseignement supérieur ou la recherche. La première moitié parle des sujets en lien avec l'enseignement supérieur et la seconde moitié des sujets en lien avec la recherche.

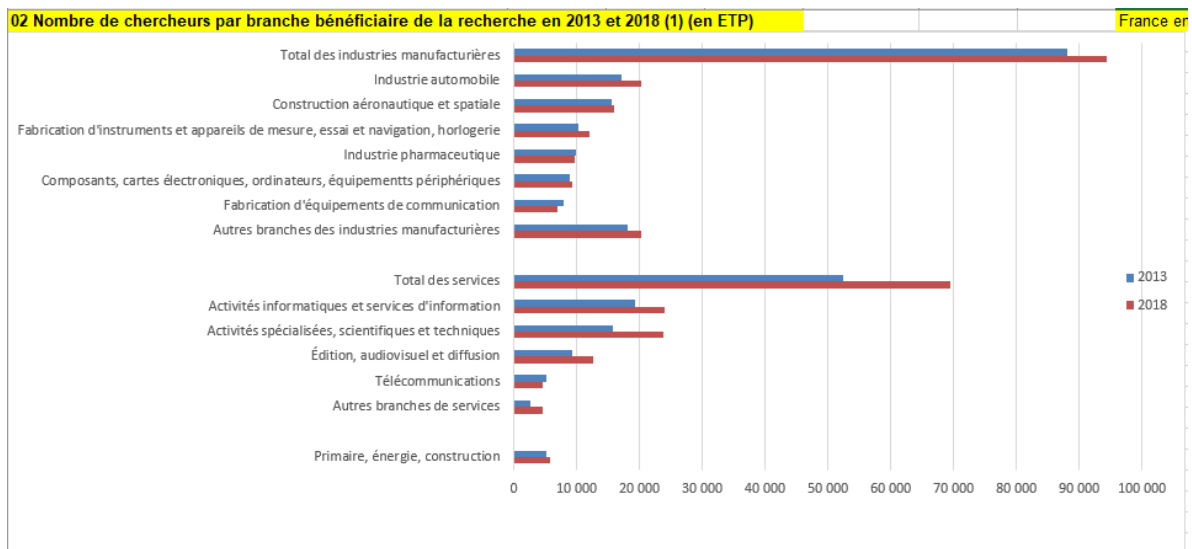
Pour la réalisation de cette brochure, la majeure partie des données sont fournies par le SSM. Mais, elle s'appuie également sur les autres sources de données, en particulier celles provenant de l'Insee, du MESRI et de l'OCDE.

J'ai commencé cette mission le 15 Octobre 2020 et elle s'arrête le 16 Novembre 2020. C'est une mission annuelle. Je suis chargé de faire la synthèse et de produire des illustrations graphiques sur les fiches 35 et 36. La fiche 35 aborde le sujet des moyens humains de la recherche et développement. La fiche 36 parle de la parité dans la recherche.

3.1.1 La fiche 35 : les moyens humains de la R&D

Dans la fiche 35, on comptabilise le nombre d'actifs en lien avec une activité de R&D en France et on mesure leur taux d'accroissement par rapport aux années précédentes en ETP (Equivalent temps plein) ou en PP (Personne physique). On distingue les chercheurs dans le domaine privé et dans le domaine public.

Dans le secteur public, on compte les chercheurs qui travaillent pour l'Etat, les chercheurs dans l'enseignement supérieur et les chercheurs dans les ISBL. Parmi les chercheurs dans le secteur privé, ils sont répartis sur les trente-deux branches d'activités (NAF rév2). Ensuite, on étudie la part des femmes parmi les chercheurs dans le secteur privé et public. On compare également la France avec les autres pays de l'OCDE sur le nombre de chercheurs en proportion de population active.



Le nombre de chercheur en ETP par branche bénéficiaire de la recherche en 2013 et 2018

Voici un graphique qui illustre le nombre de chercheurs en ETP dans les trente-deux branches d'activités en France en 2013 (en bleu) et en 2018 (en rouge).

Sur ce diagramme en barre groupé, j'ai représenté les secteurs d'activité où le nombre de chercheurs en ETP étaient les plus importants. Ils sont classés par ordre décroissant. Nous avons dans la partie du haut le secteur secondaire (56 %), au milieu le secteur tertiaire (41 %) et en bas le secteur primaire (3 %). Avec un nombre plus important de chercheurs dans le secteur secondaire que dans les secteurs primaire et tertiaire.

Sur les trente-deux branches, quatre concentrent près de la moitié de l'effectif de chercheur en ETP : « Activités spécialisées, scientifiques et technique » (14 %), « Activités informatiques et services d'information » (14 %), « Industrie automobile » (12 %) et « Construction aéronautique et spatiale » (9 %).

Nous pouvons également constater que le nombre de chercheurs en ETP a considérablement augmenté dans le secteur tertiaire entre 2013 et 2018, soit

une augmentation (33 %) près de 5 fois plus vite que le secteur secondaire (7 %). Le secteur primaire a aussi connu une augmentation significative (10 %) entre 2013 et 2018.

Bien que le nombre de chercheurs dans le secteur tertiaire ait connu une forte croissance au cours de ces 5 années, le secondaire reste le secteur où les effectifs de chercheurs sont les plus importants.

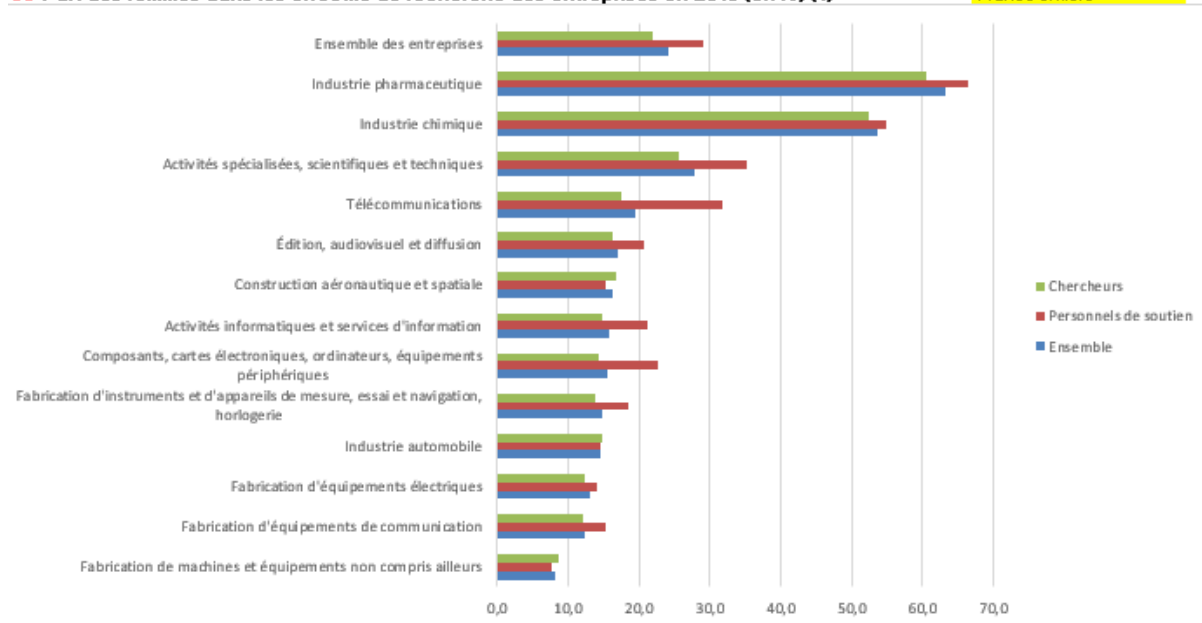
3.1.2 La fiche 36 : La parité dans la recherche

Dans la fiche 36, on mesure le nombre de femmes dans la R&D en France et son évolution. On regarde la proportion des femmes parmi les chercheurs et parmi les personnels de soutien à la recherche dans le secteur public et dans le secteur privé.

Dans le secteur privé, on regarde quelles sont les branches d'activité où les femmes sont les plus représentées dans la recherche. Dans le domaine public, on cherche également à savoir quelles sont les institutions dans lesquelles les femmes chercheuses et les personnelles de soutien à la recherche de sexe féminin sont les plus représentées. Ensuite on compare la situation de la France avec les autres pays de l'OCDE.

03 Part des femmes dans les effectifs de recherche des entreprises en 2018 (en %) (1)

France entière



La part des femmes dans les effectifs de recherche des entreprises en 2018 (en %)

Nous avons ici un diagramme en barre groupé qui illustre la part des femmes parmi les chercheurs et les personnels de soutien dans les différentes branches d'activités en 2018. Ce diagramme est trié par ordre décroissant, à part la première ligne qui représente la situation parmi l'ensemble des entreprises.

Nous avons d'un côté le pourcentage de chercheuses en vert, de l'autre les personnels de soutien en rouge et l'ensemble du personnel de recherche en bleu. Dans les entreprises, les chercheuses sont plus nombreuses que les chercheurs dans l'industrie pharmaceutique (60 %) et dans l'industrie chimique (52 %). En revanche, les femmes sont très peu représentées parmi les chercheurs dans les branches de la fabrication de machines et équipement de communication non compris ailleurs (9 %), la fabrication d'équipement de communication (12 %) et la fabrication d'équipement électrique (12 %).

Quand on regarde la part du personnel de recherche de sexe féminin (25 %) parmi l'ensemble des entreprises privées, nous sommes encore très loin pour atteindre la parité et l'égalité femme-homme.

Pour la production des tableaux et des graphiques, j'ai utilisé principalement Microsoft Excel. Les cours que j'ai eus sur Excel à l'IUT m'ont été d'une grande aide pour la réalisation de ces illustrations.

J'ai rencontré des difficultés quand j'ai commencé cette mission, elles sont principalement liées à la méconnaissance de l'environnement.



	str-dgesip-dgri-a2-1-recherche
Type :	Dossier de fichiers
Emplacement :	M:\
Taille :	133 Go (143 422 220 891 octets)
Taille sur le disque :	133 Go (143 576 342 528 octets)
Contenu :	80 429 Fichiers, 11 787 Dossiers

La base de données du département recherche

J'étais nouveau dans le département et pour produire ces deux fiches je dois savoir où aller chercher les informations. Le département dispose d'une base de données volumineuse, il y a plus de quatre-vingt mille fichiers et plus de onze mille dossiers. La grande difficulté pour moi était de savoir quelles sont les données dont j'ai besoin et dans quel répertoire je peux les trouver.

Par exemple pour la réalisation de l'illustration numéro 2 de la fiche « La parité dans la recherche ». Je dois trouver les données sur la part des femmes de

l'Institut Pasteur. Mais il n'y avait pas cette donnée, car l'Institut Pasteur est regroupé avec les autres associations (ISBL). J'ai cherché cette information dans le répertoire « Public » et dans le répertoire « Résultats généraux » sans le trouver. A la fin, j'ai obtenu ces données en demandant au responsable des enquêtes du côté administration Matthieu WYCKAERT.

Une fois que ces fiches sont terminées. Elles sont relues et validées par ma maîtresse d'apprentissage Madame Séverine MAYO-SIMBSLER. Quand toutes les fiches sont finies et validées, elles seront envoyées à la sous-directrice des SIES. Une fois que ce travail est fait, on procède à la production de la version en ligne. On travaille avec le département des outils d'aide à la décision, les Directions générales et le cabinet ministériel.

La production de la brochure EESRI est annuelle, elle sera mise en ligne au mois d'avril. La version papier sera imprimée pour le printemps 2021 et sera distribuée dans toutes les institutions de recherche et dans tous les établissements d'enseignement supérieur.

	EESRI2020 (6 mois après le lancement)	EESRI 2019 (3 mois après lancement)		Titre	Nombre de visiteurs
Préface	288	250	R	le doctorat et les docteurs	1 369
			ES	les étudiants en situation de handicap dans l'enseignement supérieur	979
Résumé	698	400	ES	l'aide sociale aux étudiants	757
			R	la dépense d'éducation pour l'enseignement supérieur	750
Enseignement supérieur	11 151	4 500		l'état de l'Enseignement supérieur, de la recherche et de l'innovation - résumé	698
			R	l'effort de recherche et développement en France	675
			R	la position scientifique de la France dans le monde à travers ses publications	629
Recherche	8 202	4 580	ES	le niveau d'études de la population et des jeunes	625
			ES	l'accès à l'enseignement supérieur	617
			ES	les bibliothèques universitaires	607

Source : [Matomo](#) – période Mai - Octobre 2020 pour l'édition 2020 et période juillet – septembre 2019 pour l'édition 2019

Source : [Matomo](#) – période Mai - Octobre 2020

Le nombre de consultation de la brochure EESRI 2020 sur le site du MESRI (Source Matomo)

Le site en ligne dédié à l'EESRI est consulté plus de vingt mille fois pour l'édition 2020 après 6 mois de lancement. Parmi ces consultations, la fiche sur le doctorat et les docteurs a eu le plus de succès (1369 visiteurs).

3.2 Intégration des données du département recherche vers l'OpenData

Le site Repères a été créé par le département des études statistiques de la recherche afin de rendre accessibles les données statistiques de la recherche aux grands publics et de publier des grands indicateurs statistiques sur la R&D en France.

Il met en ligne des fichiers PDF tels que les Notes flash et les Notes d'information. Il fournit des données statistiques sous forme de fichier Excel. Il met également à disposition des illustrations cartographiques afin d'éclaircir la situation de la recherche et innovation en France.

Les formulaires d'enquêtes sont aussi publiés sur ce site afin de faciliter le recueil des données.

La création de ce site permet aussi au département d'être indépendant de tout autre site. Ainsi, il est libre de publier les contenus qu'il lui semble nécessaire et de pouvoir répondre rapidement aux besoins des utilisateurs.

Cependant, la forme du site est très vieille, désordonnée (cf. [annexe 3](#)), ce qui le rend très difficile à utiliser. Il est donc nécessaire de procéder à une refonte.

La refonte du site est en partie faite par un apprenti (cf. [annexe 4](#)) qui était présent dans le département. Avec les nouvelles politiques gouvernementales

en matière de Web, les sites satellites sont amenés à disparaître. Donc, le site Repères intégrera le site principal du Ministère et il faut qu'il corresponde aux normes définies par ce dernier.

Ce site est hébergé par le CISAD (service informatique de la Depp, SSM du MEN), pour tout changement ou toutes modifications sur le site, on doit passer par ce service. Mais nous sommes dans l'optique d'intégrer le site principal, donc nous sommes également en collaboration avec le bureau de la communication du Ministère afin de trouver une solution pour la transition de ce site vers le site du MESRI.

Le projet est toujours en cours. Je suis chargé de retraiter les données statistiques et les fichiers afin qu'ils correspondent aux normes définies par l'OpenData et de les intégrer par la suite.

Les fichiers tels qu'ils sont sur le site Repères ne sont pas exploitables par les utilisateurs. Certaines données sont également erronées et nécessitent une correction.

Pour cette mission, je travaille principalement avec Thomas BALCONE, responsable des enquêtes auprès des entreprises et le département des outils d'aide à la décision. C'est lui qui supervise mes travaux.

La principale difficulté est de savoir comment mettre en forme ces données. Certains fichiers rassemblent plusieurs variables de nature différentes, faut-il créer plusieurs fichiers et de les répartir par leurs natures ? Il faut également savoir quelles sont les données à mettre en avant, quelles sont les données non utilisables.

En millions d'euros	(4)	
	1991	1992
Chercheurs en ETP (1)	59,594	64,688
Effectifs totaux de R&D en ETP (1)	156,300	164,378
Financement de la DIRDE	15,286	16,134
- par les entreprises	10,266	11,222
- par l'Etat (2)	3,271	2,972
- par l'étranger	1,750	1,940
Dépenses extérieures exécutées (3)	3,406	3,717
- par les entreprises en France	2,125	2,187
- par l'Etat	284	333
- par l'étranger	997	1,197

Fichier ent1 original

Voici un exemple de la mise en forme des fichiers. L'image ci-dessus correspond au fichier original « ent1 », dans ce fichier, nous avons d'un côté des données sur les effectifs de chercheurs en ETP et de l'autre des données sur le financement de la DIRDE. L'unité de compte (ici en euro) n'est pas vraiment adéquate pour compter des personnes.

Ce sont donc des données de nature complètement différente. Après réflexion, j'ai séparé le fichier en deux. J'ai créé un fichier sur les chercheurs et un autre sur le financement.

Année	Indicateur	Code	Déclinaison	Valeur
1991	Dépenses totales de R&D des entreprises	DIRDE	Dépenses intérieures de R&D des entreprises	15,286.22
1991	Dépenses totales de R&D des entreprises	DIRDE_FIN_ENT	Financées par les entreprises implantées en France	10,265.61
1991	Dépenses totales de R&D des entreprises	DIRDE_FIN_ETAT	Financées par l'Etat	3,271.10
1991	Dépenses totales de R&D des entreprises	DIRDE_FIN_ETRANGER	Financées par l'Etranger	1,749.50
1991	Dépenses totales de R&D des entreprises	DERDE	Dépenses extérieures de R&D des entreprises	3,406.32
1991	Dépenses totales de R&D des entreprises	DERDE_ENT	DERDE vers les entreprises implantées en France	2,125.14
1991	Dépenses totales de R&D des entreprises	DERDE_ETAT	DERDE vers l'Etat	283.86
1991	Dépenses totales de R&D des entreprises	DERDE_ETR	DERDE vers l'Etranger	997.32
1991	Dépenses totales de R&D des entreprises	DEP_TOT_ENT_RD	Dépenses totales de R&D des entreprises	16,567.40
1991	Dépenses totales de R&D des entreprises	DEP_TOT_ENT_RD_FIN_ENT	Financées par les entreprises implantées en France	11,394.96
1991	Dépenses totales de R&D des entreprises	AUTOFINANCEMENT	Autofinancement	9,339.95
1991	Dépenses totales de R&D des entreprises	FINANCEMENT_PAR_AUTRES_ENT	Financées par les autres entreprises	2,055.01
1991	Dépenses totales de R&D des entreprises	DEP_TOT_ENT_RD_FIN_ETAT	Financées par l'Etat	3,422.94

Fichier ent1 après la mise en forme - données sur les dépenses et financements des entreprises

Année	Indicateur	Code	Taille entreprise	Déclinaison	Valeur
1991	Effectifs de R&D des entreprises en ETP	EFFRD_ETP		Effectifs R&D totaux des entreprises en ETP	156,300.00
1991	Effectifs de R&D des entreprises en ETP	CHERCHEUR_ETP		Chercheurs et ingénieurs de R&D (y compris doctorants) en ETP	59,594.00
1991	Effectifs de R&D des entreprises en ETP	TECH_ET_SUPPORT_RD_ETP		Techniciens et personnels de support R&D en ETP	96,706.00
1992	Effectifs de R&D des entreprises en ETP	EFFRD_ETP		Effectifs R&D totaux des entreprises en ETP	164,378.00
1992	Effectifs de R&D des entreprises en ETP	CHERCHEUR_ETP		Chercheurs et ingénieurs de R&D (y compris doctorants) en ETP	64,688.00
1992	Effectifs de R&D des entreprises en ETP	TECH_ET_SUPPORT_RD_ETP		Techniciens et personnels de support R&D en ETP	99,690.00
1993	Effectifs de R&D des entreprises en ETP	EFFRD_ETP		Effectifs R&D totaux des entreprises en ETP	164,384.00
1993	Effectifs de R&D des entreprises en ETP	CHERCHEUR_ETP		Chercheurs et ingénieurs de R&D (y compris doctorants) en ETP	66,455.00
1993	Effectifs de R&D des entreprises en ETP	TECH_ET_SUPPORT_RD_ETP		Techniciens et personnels de support R&D en ETP	97,929.00
1994	Effectifs de R&D des entreprises en ETP	EFFRD_ETP		Effectifs R&D totaux des entreprises en ETP	161,955.00
1994	Effectifs de R&D des entreprises en ETP	CHERCHEUR_ETP		Chercheurs et ingénieurs de R&D (y compris doctorants) en ETP	66,714.00
1994	Effectifs de R&D des entreprises en ETP	TECH_ET_SUPPORT_RD_ETP		Techniciens et personnels de support R&D en ETP	95,241.00

Fichier ent1 après la mise en forme - données sur les effectifs des personnels de recherche (en ETP) en entreprise

Ensuite, il faut définir la nomenclature. Par exemple pour le fichier sur les effectifs de chercheurs, nous avons les données sur l'effectif total de personnels

de R&D (EFFRD_ETP) et les données sur l'effectif des chercheurs (CHERCHEUR_ETP). Mais il manque les données sur les personnels de soutien à la recherche. C'est à moi de choisir un code qui soit le plus intuitif possible pour représenter cette donnée, et elle s'obtient par la formule suivante :

$$TECH_ET_SUPPORT_RD_ETP = EFFRD_ETP - CHERCHEUR_ETP$$

Une fois ces travaux faits, ces fichiers sont ensuite vérifiés par Thomas et une fois validés, il les transmet à l'adjoint au chef du département des outils d'aide à la décision et les fichiers seront mis en ligne sur OpenData.

J'ai transformé au total 44 fichiers, chaque fichier comporte entre 500 et 3000 lignes de données.

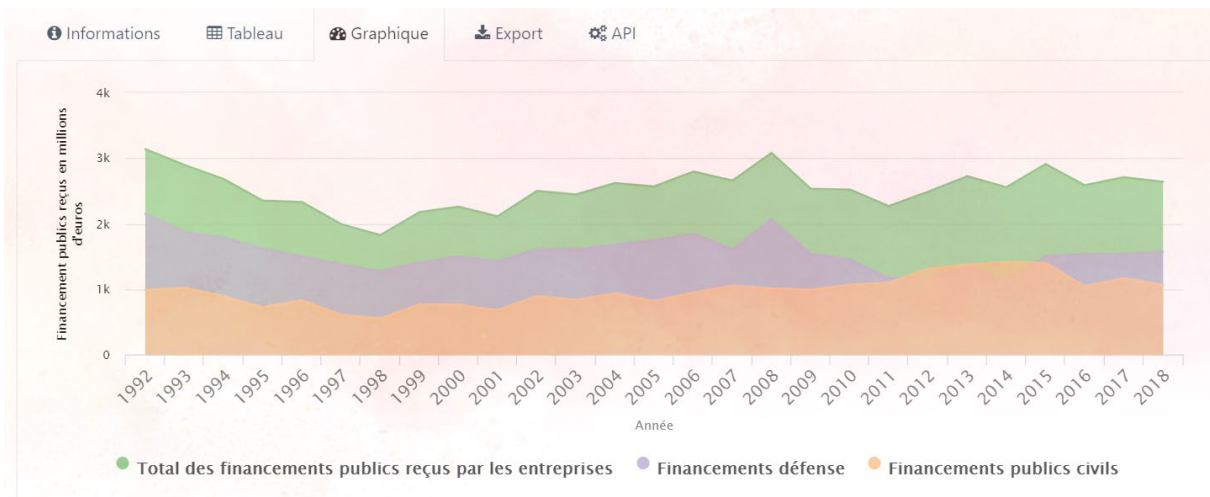
Financement publics reçus par les entreprises (en millions d'euros)

Année	Code indicateur	Indicateur	Code déclinaison	Déclinaison	Valeur	
1	2018	indicateur_FPE	Les financement publics re...	FPE_0113	CNES	117,38
2	2018	indicateur_FPE	Les financement publics re...	FPE_0123	Ministère en charge de l'e...	54,54
3	2018	indicateur_FPE	Les financement publics re...	FPE_0114	Autres grands programme...	9,34
4	2018	indicateur_FPE	Les financement publics re...	FPE_0122	Ministère de l'industrie (h...	344,88
5	2018	indicateur_FPE	Les financement publics re...	FPE_0124	Autres	313,53
6	2018	indicateur_FPE	Les financement publics re...	FPE_0111	Direction des programmes...	24,77
7	2018	indicateur_FPE	Les financement publics re...	FPE_0112	Ministère en charge de l'in...	29,09
8	2018	indicateur_FPE	Les financement publics re...	FPE_012	Ministères autres que le m...	782,11
9	2018	indicateur_FPE	Les financement publics re...	FPE_013	Autres financements publi...	102,08
10	2018	indicateur_FPE	Les financement publics re...	FPE_01	Financements publics civils	1 064,78
11	2018	indicateur_FPE	Les financement publics re...	FPE_02	Financements défense	1 572,43

Partager Intégrer Widget

<https://data.enseignementsup-recherche.gouv.fr/explore/embed/dataset/fr-esr-financement-publics-recus-par-les-entreprises/table/?disjunctive.c>

Fichier sur le financement public reçus par les entreprises disponibles sur le site OpenData



Visualisation graphique des données sur le financement public reçu par les entreprises entre 1992 et 2018

3.3 Les réponses à la demande

Le département des études statistiques de la recherche reçoit tous les ans des demandes qui proviennent de tout horizon (cabinet ministériel, Insee, collectivité territoriale, bureau de conseil, institut de recherche, ...). La majorité des demandes qui nous sont adressées réclament des données et des agrégats statistiques sur la recherche en France, que ce soit dans les administrations ou dans les entreprises.

Certaines demandes concernent également des précisions (notamment les secrets statistiques sur les associations et les GIP) sur les enquêtes annuelles réalisées par notre département.

Dans les années précédentes, les demandes sont directement adressées aux chargés d'études ou à un membre du département concerné. Cela ne facilite pas le suivi des demandes, on ne peut pas savoir si les demandes sont traitées dans les délais impartis ou non. De plus il y a une mauvaise organisation dans la gestion des fichiers et des répertoires concernant les demandes (plusieurs répertoires, sauvegarder dans des endroits différents, ...). Ce n'est pas facile de retrouver des éléments de réponses si on veut revenir sur une ancienne demande.

Conscient et soucieux de ces problèmes, on a décidé avec la cheffe de département, son adjointe, la secrétaire et moi de centraliser ces demandes et d'organiser des réunions de suivi hebdomadaires afin d'assurer le bon déroulement du traitement des demandes.

Nous avons d'abord créé un répertoire « Demandes » dans l'espace partagé afin que tout le monde puisse avoir accès et tous les éléments de réponses sont

également stockés dans ce répertoire avec un répertoire pour chaque demande. Ensuite toutes les demandes passent désormais par la boîte fonctionnelle du département, après en fonction de l'objet de la demande nous l'adressons aux personnes compétentes qui sont capables de fournir des éléments de réponses.

Nous partageons un tableur Excel de suivi des demandes avec le département des études statistiques de l'enseignement supérieur. Sur ce tableau, nous mettons des informations telles que le numéro de la demande, l'objet de la demande, la personne ou l'institution qui nous a adressé cette demande, les instructeurs ainsi que la date de livraison.

N° Ordre	Date arrivée SIES	Date de livraison attendue	Département concerné	Qui	Contact	Organisme demandeur	Typologie
-------------	----------------------	----------------------------------	-------------------------	-----	---------	---------------------	-----------

Une partie de l'en-tête du tableau de suivi

J'ai répondu à deux demandes qui nous sont adressées. La première demande concerne les données de dépenses en R&D dans les Dom Tom. C'est une personne travaillant dans un bureau de conseils qui nous a adressé cette demande. A partir des données dont nous disposons, j'ai fourni à cette personne des données sur les dépenses intérieures de R&D, des données sur les effectifs en ETP et des données sur les chercheurs en ETP également.

La seconde demande à laquelle j'ai répondu est toujours une demande des données, mais cette fois-ci sur les chercheurs de la région Bourgogne-Franche - Comté. Cette demande vient du conseil régional de Bourgogne-Franche-Comté. J'ai fait la synthèse des données et des informations qui m'ont été fournies.

Certaines données sont trouvables sur Eurostat et le reste on les a complétés avec la base de données du département.

Grâce à la mise en place de ces mesures de suivi, cette année nous avons répondu à 39 demandes

3.4 Le Funathon

Un Funathon est un évènement de formation collaboratif qui nous offre des opportunités pour s'entraîner sur des techniques de Data Science. Cet évènement est organisé par le SSP Lab de l'INSEE, la durée est de deux jours.

Le SSP Lab est une entité au sein de la division des Méthodes Appliquées de l'Econométrie (DMCSI) créée en 2018. On compte parmi cette entité des Data Scientistes et des économètres. Ils ont pour mission de stimuler l'innovation et la R&D dans la production des statistiques avec les nouvelles méthodes des science des données. D'autres parts, ils organisent des évènements et des travaux collaboratifs avec les Services Statistiques Ministérielles (SSM) afin d'assurer une veille technologique et de diffuser les nouvelles techniques de traitement et de valorisation des données à différents niveaux.

Pour participer à cet évènement, il faut déjà être familier avec un langage de programmation tel que Python ou R. Le but de ce Funathon est de familiariser les agents avec les techniques de data science. Pour les personnes plus expérimentées, cette formation permet également de se perfectionner et de consolider les acquis.

J'ai formé une équipe avec une personne qui vient de l'OCDE et une autre personne issue de l'INSEE afin de faciliter l'échange et le travail collaboratif.

Le sujet étudié est celui de l'immobilier, en l'occurrence ici Le SSP Lab nous a mis à disposition des données issues d'Airbnb sous forme de fichier csv. Ce sont les données qui présentent l'ensemble des locations réalisées via Airbnb dans les trois grandes villes de France (Paris, Lyon, Bordeaux). Le thème de l'immobilier est intéressant, les données fournies sont de nature différente, ce qui permet d'aborder un large spectre de domaine en data science tels que : le machine learning, analyse textuelle, le nlp (natural language processing), dataviz, ...

3.4.1 Présentation des données

Nous avons à notre disposition les données qui proviennent d'Inside Airbnb. Les données sur ce site sont issues d'une démarche citoyenne qui rend accessible les données sur Airbnb, les données ont été nettoyés et agrégés pour alimenter le débat public.

Nous avons principalement trois types de données :

- La liste des logements
- La liste des commentaires
- Le détail du calendrier

Les données sur les logements contiennent des informations détaillées sur tous les logements : descriptions textuelles, tous les champs de description et l'ajout des informations complémentaires par le site Airbnb.

```
[8]: listings.head()
```

[8]:	id	listing_url	scrape_id	last_scraped	name	description	neighborhood_overview
0	2577	https://www.airbnb.com/rooms/2577	20210208234805	2021-02-14	Loft for 4 by Canal Saint Martin	100 m2 loft (1100 sq feet) with high ceiling, ...	NaN https://a0.mu
1	3109	https://www.airbnb.com/rooms/3109	20210208234805	2021-02-13	zen and calm	The space <l bedroom apartment in ...	Good restaurants <very close the Montparna...
2	5396	https://www.airbnb.com/rooms/5396	20210208234805	2021-02-12	Explore the heart of old Paris	Cozy, well-appointed and graciously designed s...	You are within walking distance to the Louvre,...
3	7397	https://www.airbnb.com/rooms/7397	20210208234805	2021-02-14	MARAIS - 2ROOMS APT - 2/4 PEOPLE	VERY CONVENIENT, WITH THE BEST LOCATION ! <br /...>	NaN https://a0.mu
4	7964	https://www.airbnb.com/rooms/7964	20210208234805	2021-02-13	Large & sunny flat with balcony !	Very large & nice apartment all for you! <br /...>	NaN https://a0.m

5 rows x 74 columns

Le début des données sur les logements

Les données sur les commentaires sont présentes uniquement sur les bases de données où le client a donné un commentaire sur le logement.

```
listings.head()
```

	listing_id	id	date	reviewer_id	reviewer_name	comments
0	2577	366217274	2019-01-02	28047930	Kate	Beautiful apartment in a really handy location...
1	3109	123127969	2016-12-27	12389804	Sophie	The host canceled this reservation the day bef...
2	3109	123274144	2016-12-28	67553494	Tom'S	The host canceled this reservation 2 days befo...
3	3109	207127433	2017-10-28	51636494	Patricia	Tout s'est bien déroulé. Merci bien. PG
4	3109	208779822	2017-11-03	4142888	Patricia	Un petit nid fouiller douillet situé dans app...

Le début des données sur les commentaires

Les données sur les périodes de location sont présentes sur le fichier calendar.csv, ces données nous permettent de connaître la réservation ou non de la période concernée.

listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
77482	7/7/2021	f	\$400.00	\$400.00	3	60
2577	7/8/2021	t	\$125.00	\$125.00	3	1125
2577	7/9/2021	t	\$125.00	\$125.00	3	1125
2577	7/10/2021	t	\$125.00	\$125.00	3	1125
2577	7/11/2021	t	\$125.00	\$125.00	3	1125
2577	7/12/2021	t	\$125.00	\$125.00	3	1125
2577	7/13/2021	t	\$125.00	\$125.00	3	1125
2577	7/14/2021	t	\$125.00	\$125.00	3	1125
2577	7/15/2021	t	\$125.00	\$125.00	3	1125
2577	7/16/2021	t	\$125.00	\$125.00	3	1125
2577	7/17/2021	t	\$125.00	\$125.00	3	1125
2577	7/18/2021	t	\$125.00	\$125.00	3	1125
2577	7/19/2021	t	\$125.00	\$125.00	3	1125
2577	7/20/2021	t	\$125.00	\$125.00	3	1125
2577	7/21/2021	t	\$125.00	\$125.00	3	1125
2577	7/22/2021	t	\$125.00	\$125.00	3	1125

Une partie des données sur les réservations (données temporelles)

Ces données sélectionnées ne concernent seulement les trois villes de France : Paris, Bordeaux et Lyon.

La problématique que nous nous sommes posés pour ce Funathon est la suivante :

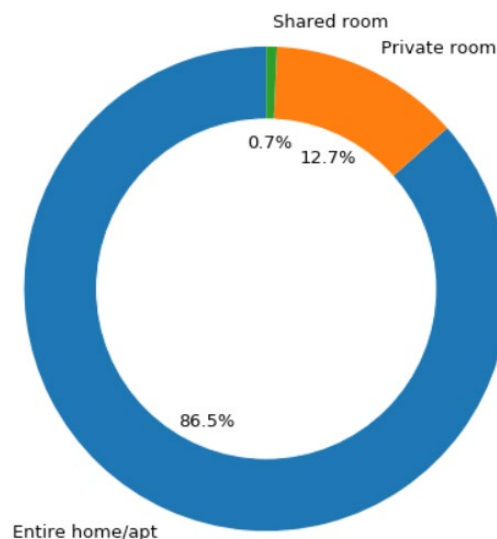
Le droit à la ville est-il menacé par Airbnb ?

En effet, les plateformes numériques telles que Airbnb, Uber, sont en train d'envahir le monde, nous sommes de plus en plus habitués à utiliser leurs applications et services. Ce phénomène intéresse et inquiète les acteurs publics, la prise de décisions est difficile. Choisir entre l'attractivité territoriale et l'ouverture sur le monde, ou de garantir le bien-être de leur citoyen ?

Les data scientistes, les économistes, les sociologues utilisent ces données pour comprendre les évolutions et les transformations qui sont en cours, notamment pour mesurer l'impact d'Airbnb sur le secteur de l'hôtellerie et du tourisme, et d'analyser les conflits d'usages entre les fonctions résidentielles et touristiques des logements.

3.4.2 Analyse

Nous nous sommes d'abord intéressés à quoi ressemble un logement Airbnb à Paris. J'ai fait de la statistique descriptive pour décrire les appartements loués, leurs caractéristiques. J'ai ensuite utilisé une librairie (le WordCloud) en Python pour analyser les données textuelles, d'en extraire les mots qui reviennent le plus souvent et de les transformer en nuage mots.



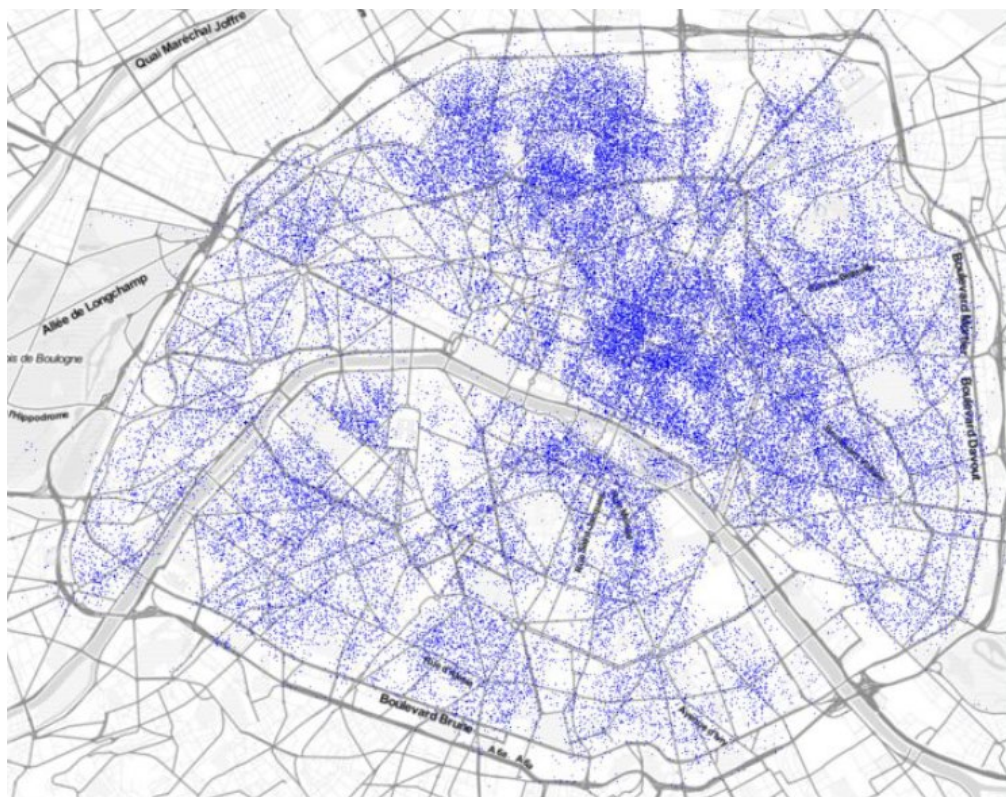
Le type de logement en %

Donc d'après ce graphique en anneau, 86,5 % des biens loués sur Airbnb sont des appartements ou des studios en entier, 12,7 % des biens loués sont des

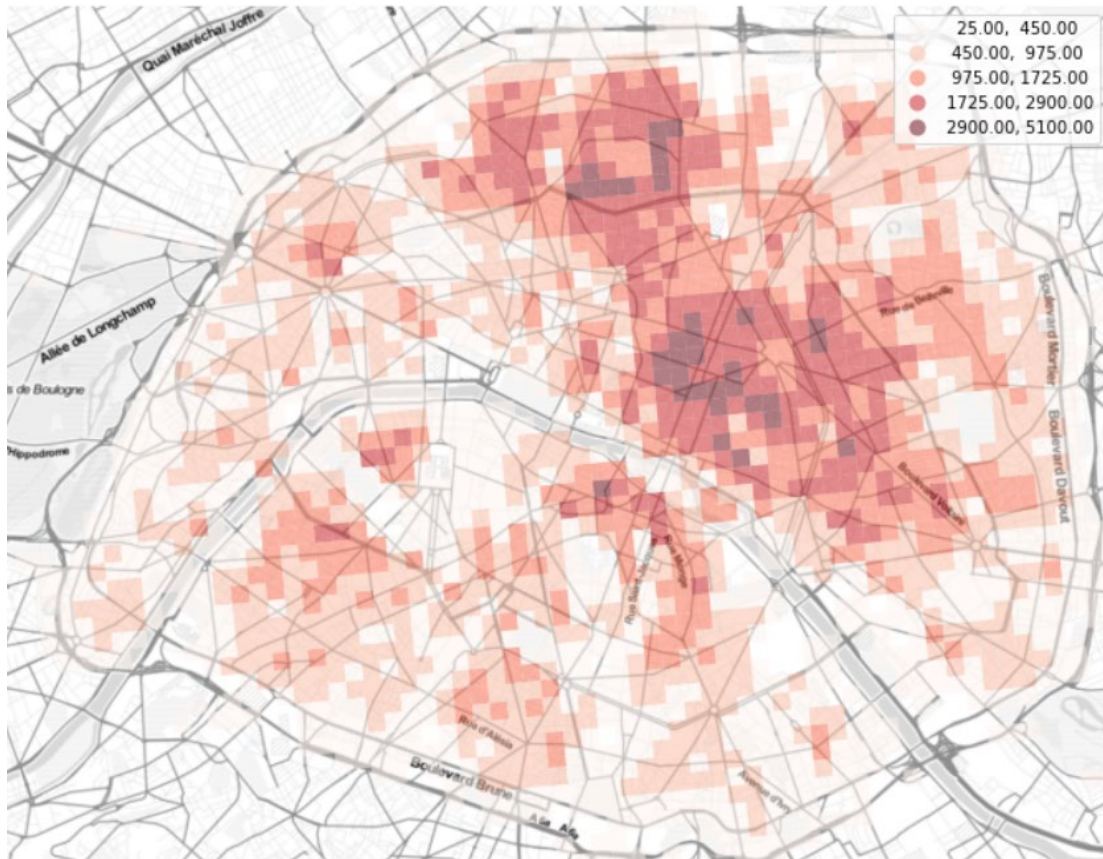


Le nombre de logement Airbnb à Paris

Il y a au total 64 293 logements Airbnb à Paris.



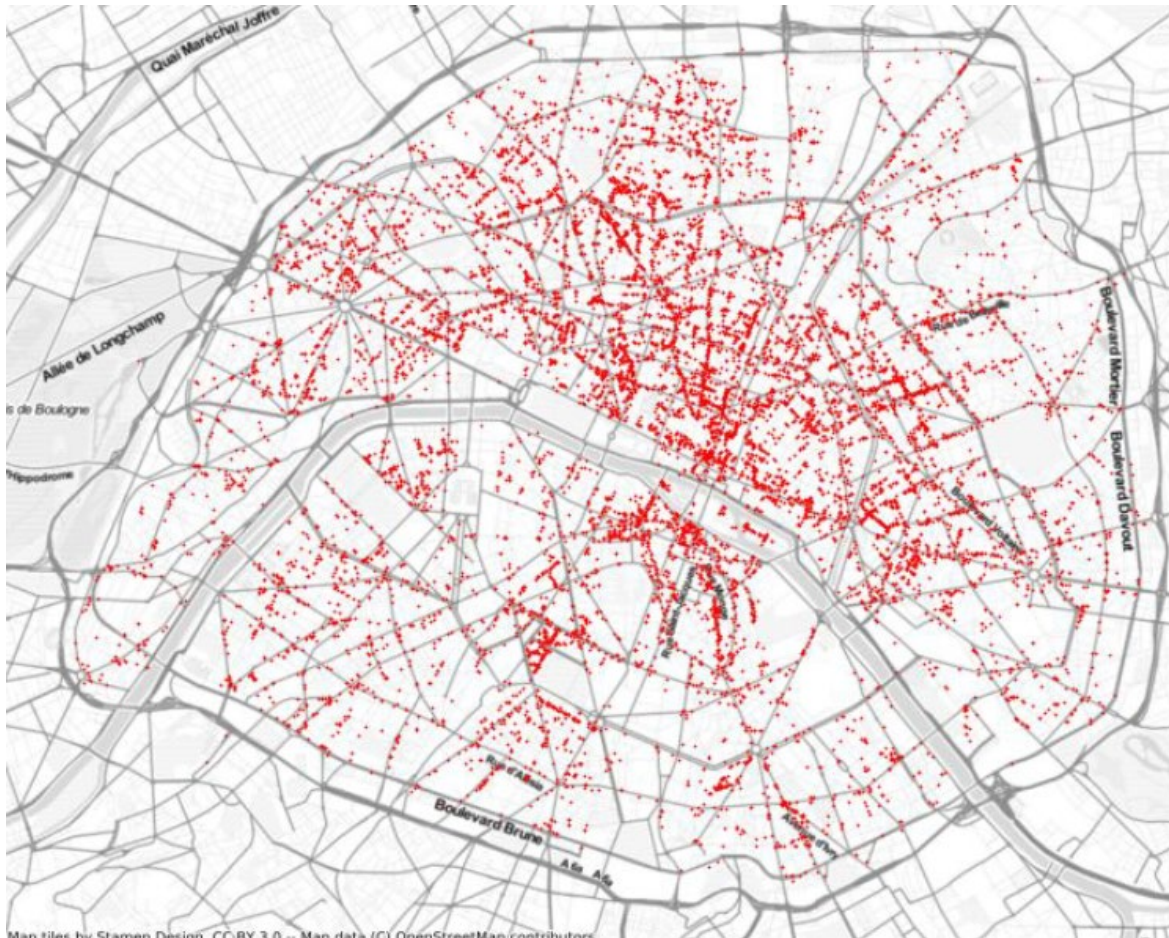
Emplacements de logements Airbnb à Paris



La densité des logements Airbnb à Paris

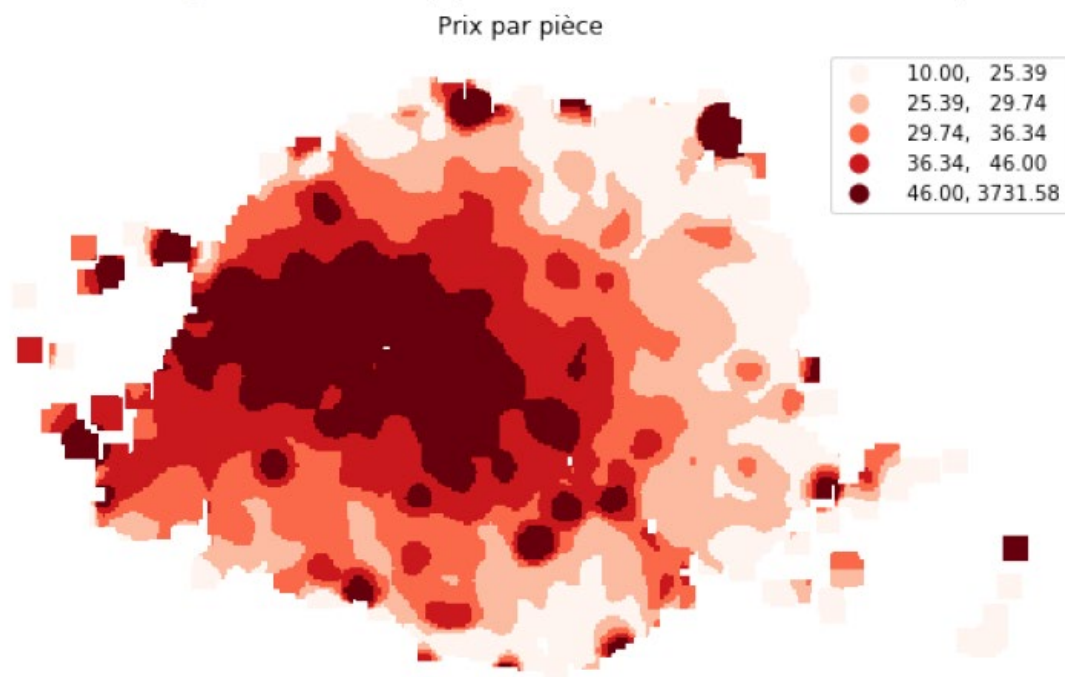
Avec la carte des emplacements des logements et la carte de densité, nous pouvons constater que les biens loués sur Airbnb se concentrent sur le centre, le nord et le nord-ouest de Paris. Dans le centre et au nord, il y a la plus grande concentration des biens (entre 2 900 et 5 100 de logements Airbnb par pixel).

Pourquoi y-a-t-il une concentration au centre et au nord de Paris ?

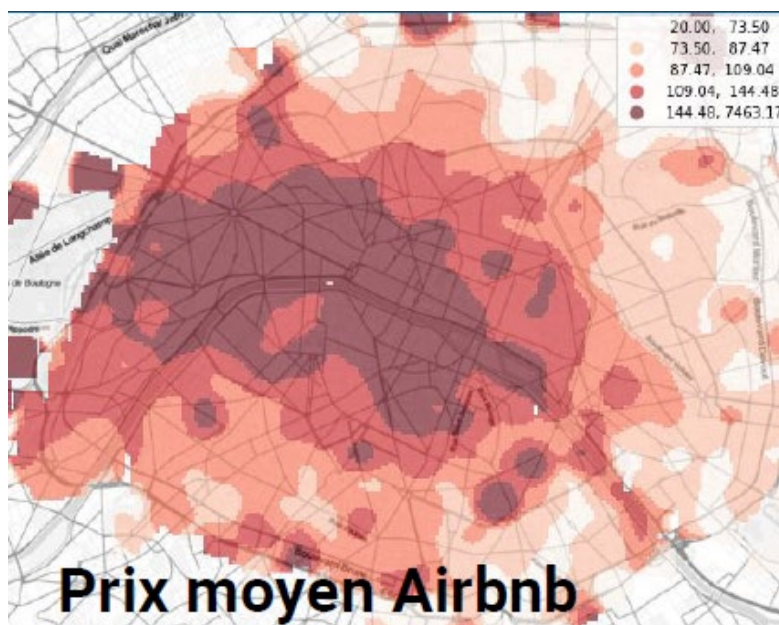


Emplacement des terrasses éphémères à Paris

Sur cette carte, nous avons les emplacements des terrasses éphémères à Paris, la densité des biens loués est en correspondance avec la densité des terrasses à Paris. Donc, sur les quartiers où il y a le plus de terrasses, il est plus facile de trouver un logement sur Airbnb.



Prix des logements rapporté au nombre de pièce à Paris



Prix moyen des logements Airbnb à Paris

Un autre facteur qui peut influencer, c'est le prix. Nous avons ici deux hit maps, le premier décrit le prix des logements rapportés au nombre de pièces et le second décrit le prix moyen d'une nuit dans un logement Airbnb. Plus le prix

est élevé, plus la couleur est foncée. Nous pouvons constater qu'au centre et à l'ouest de Paris, les prix des logements sont les plus élevés. Inversement, à l'est et au nord, les prix sont moins élevés. Donc, il y a plus de concentration de logements Airbnb au nord de Paris, car les prix des locations sont moins chers.

Mais comment expliquer qu'il y a un cluster au centre et pas à l'est ?



Carte touristique de Paris (source : google)

D'après cette carte touristique, l'emplacement du logement est important. Les logements qui se situent au centre de Paris sont à proximité des lieux touristiques fréquentés par les touristes. Ce qui explique que, même si les prix de la location d'un logement au centre est bien plus élevé qu'à l'est de Paris, les consommateurs sont quand même prêts à payer, car le trajet pour se rendre d'un lieu touristique à un autre est plus court.

Nous arrivons maintenant sur l'analyse des profils des loueurs afin d'estimer le genre et l'âge à partir d'image qui ont été fournis sur le site Airbnb. Et inversement de construire des images à partir des données textuelles. Pour faire cette étude, nous avons utilisé le Framework Deepface de Python. Cette étude n'a pas été réalisée dans son intégralité, faute de temps. Mais j'ai quand même eu une initialisation à la reconnaissance d'image.

On importe dans un premier temps les bibliothèques nécessaires :

```
import requests
from io import BytesIO
import urllib.request

import os

import s3fs

from deepface import DeepFace
import pandas as pd
```

Ensuite, on charge les modèles qui sont pré-entraînés pour la détection des visages, d'âge et de genre sur le site de SPP Lab.

```
urllib.request.urlretrieve('https://minio.lab.sspcloud.fr/
urllib.request.urlretrieve('https://minio.lab.sspcloud.fr/
urllib.request.urlretrieve('https://minio.lab.sspcloud.fr/
```

Après, on charge les données et on construit la base de données sur les loueurs à partir des fichiers csv.


```

hosts = listings_d.groupby(['host_id', 'host_url',
                           'host_name', 'host_since', 'host_location', 'host_about',
                           'host_response_time', 'host_response_rate', 'host_acceptance_rate',
                           'host_is_superhost', 'host_thumbnail_url', 'host_picture_url',
                           'host_neighbourhood', 'host_listings_count',
                           'host_total_listings_count', 'host_verifications',
                           'host_has_profile_pic', 'host_identity_verified'], as_index=False)['id'].count()

```

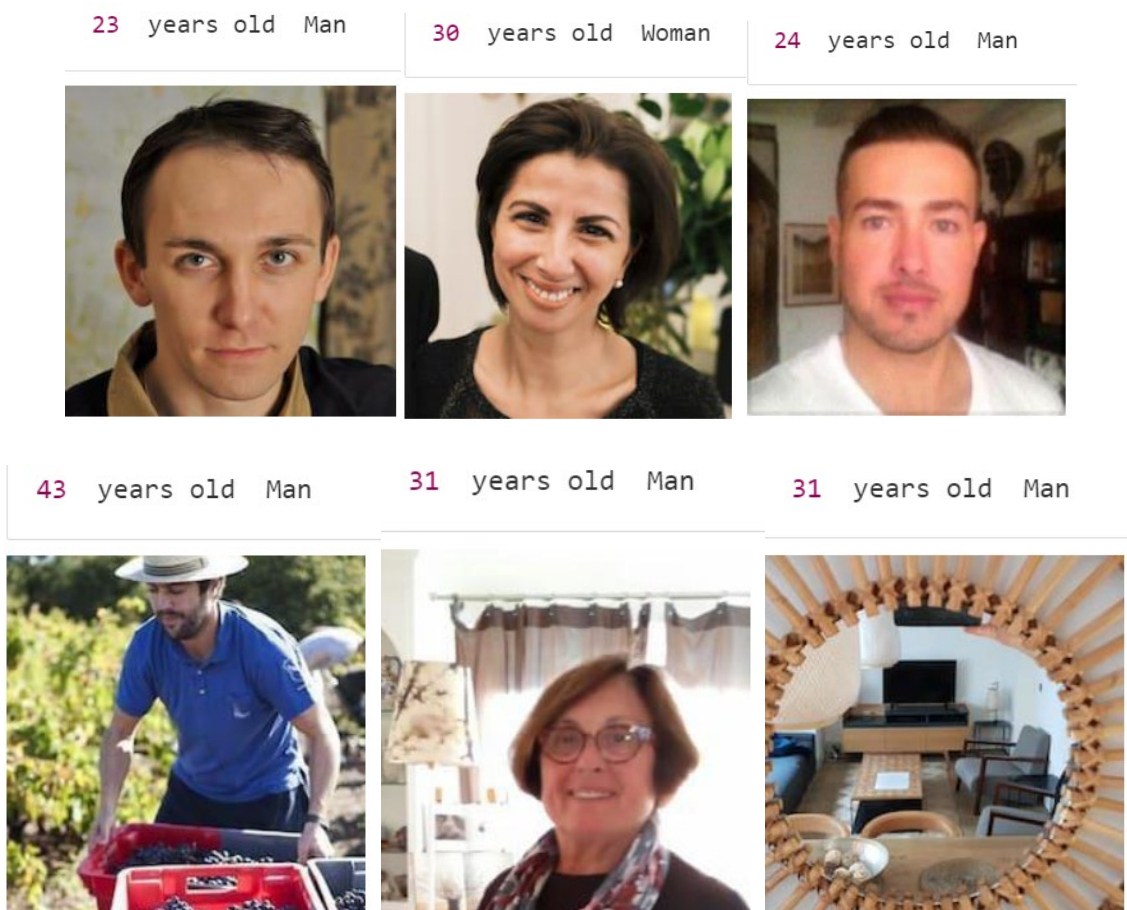
A partir de cette base, on applique les modèles pré-entraînés et on effectue les reconnaissances de l'âge et du genre

```

list_objs = []
for file in list_images:
    list_objs.append(DeepFace.analyze(img_path = file, actions = ['age', 'gender'], enforce_detection=False))

```

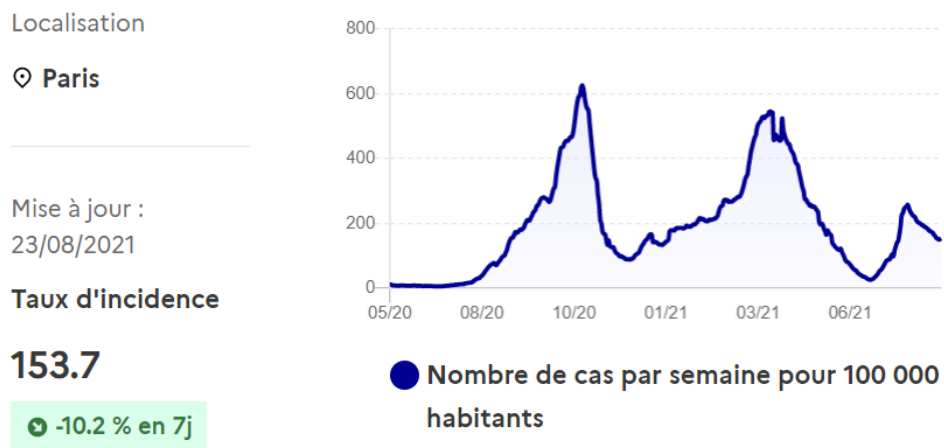
Voici quelques résultats :



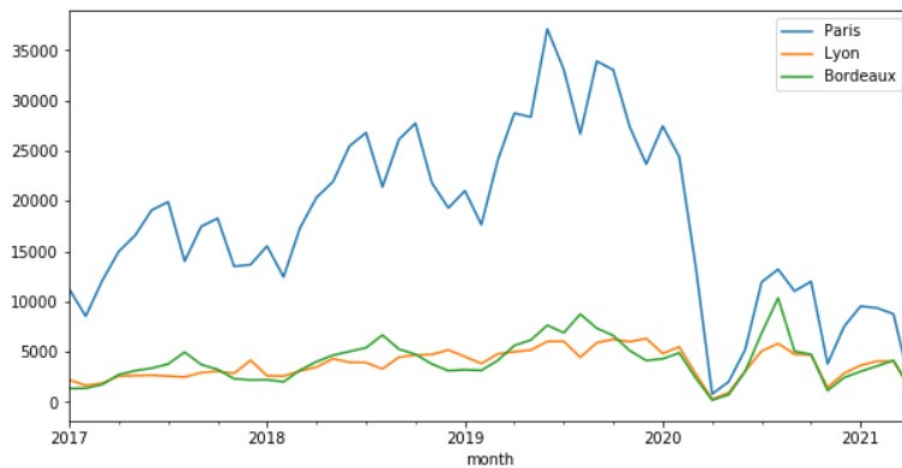
Les résultats finaux sont loin d'être parfait, il y a même beaucoup d'erreurs de reconnaissance. Mais, cette première expérience sur la reconnaissance d'image

m'a permis de voir ce qu'on est capable de faire avec les données et le machine learning. La reconnaissance d'image est un sujet très intéressant et sensible.

Nous pouvons voir également l'impacte de la crise sanitaire sur l'activité touristique et sur la réservation des logements Airbnb. J'ai fait cette analyse avec la librairie « Pandas » de Python.



L'évolution de la pandémie à Paris (source : data.gouv)



Le nombre de commentaire laissé sur Airbnb dans les trois villes

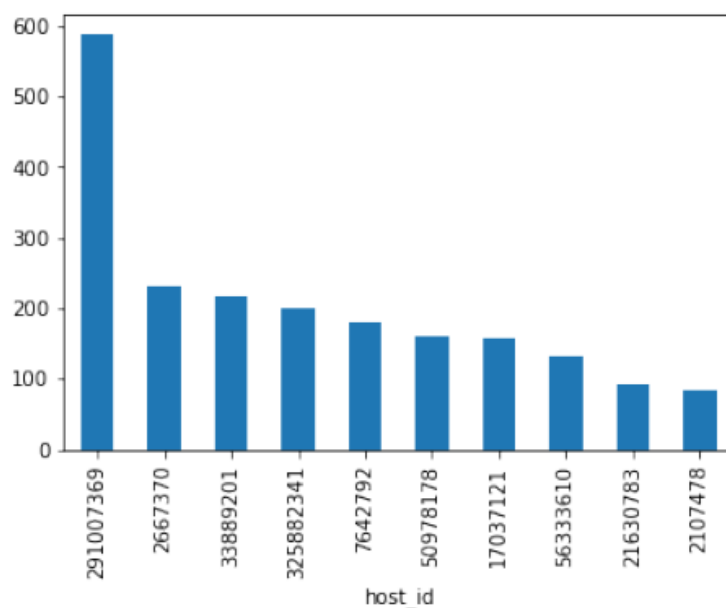
Sur le premier graphique, nous avons l'évolution du nombre de cas covid en France depuis mai 2020, sur le second nous avons le nombre de commentaires laissés par les loueurs depuis 2017, nous avons en moyenne un commentaire pour

deux réservations, l'idéal serait d'utiliser les données sur le nombre de réservations, mais cette donnée n'est pas disponible.

Donc nous pouvons voir qu'en mars 2020, le nombre de commentaires laissé a connu une chute phénoménale (presque égale à 0), car il y a eu un confinement total et les vols sont arrêtés. Les commentaires commencent à augmenter petit à petit avec l'assouplissement des mesures sanitaires, en même temps le nombre de cas contaminés augmente aussi. On arrive au novembre 2020, il y a une seconde chute du nombre de commentaires, car il y a un endurcissement des mesures sanitaires parce que le nombre de cas malade est de nouveau devenu très important.

Il y a donc une corrélation entre l'évolution de la crise sanitaire et le nombre de commentaires laissé, ce qui signifie que si le nombre de commentaires augmente c'est qu'il y a un assouplissement des restrictions et que les voyages touristiques sont de nouveaux autorisés et inversement, si le nombre de commentaires chute, c'est que les restrictions sont endurcies et on ne peut plus voyager.

Airbnb, c'est plus de 64 000 logements rien que sur Paris. Le principal hôte a presque 600 logements, c'est la société de conciergerie Luckey. La majorité des logements sont gérés par les loueurs pro.

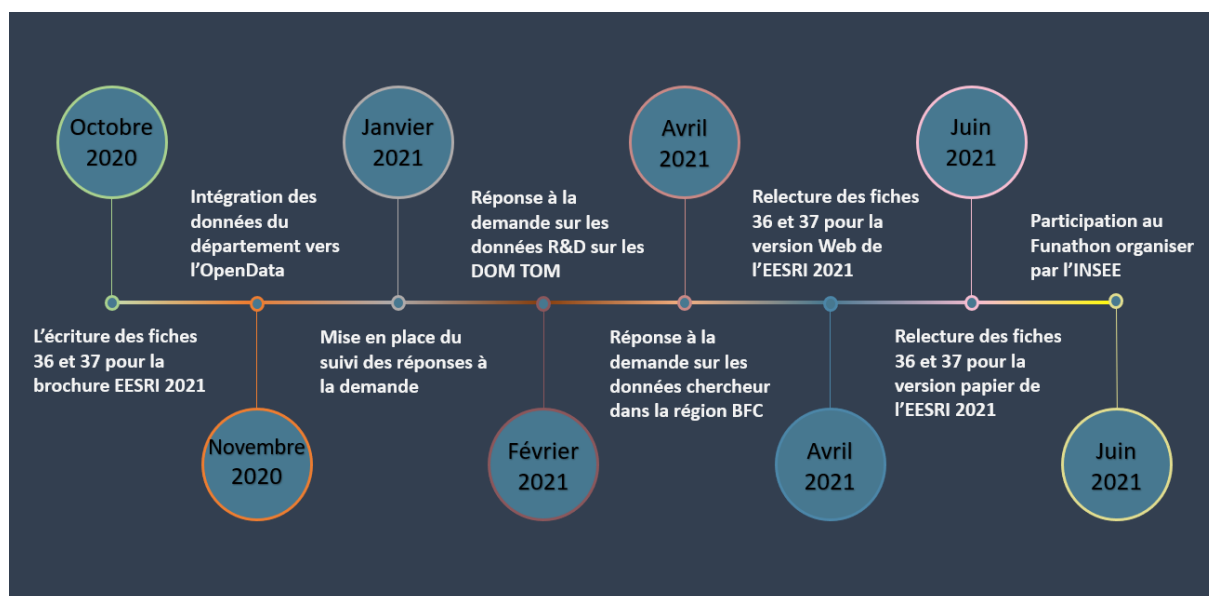


Le nombre de logements possédé par hôte

D'après une étude citoyenne publiée sur le site <https://parisvsbnb.fr/>, « depuis 2011 Paris perd plus de 11 000 habitants par an (source INSEE), plus de 29 000 logements ont été retirés du marché traditionnel locatif fin 2016 et plus de 48 000 habitants ne peuvent plus se loger ».

Airbnb est devenu le premier hôtelier de France. Le droit à la ville est bien menacé.

Le Bilan



Cette année j'ai réalisé au total 7 missions. Deux de ces sept missions m'ont occupées tout au long de l'année : l'intégration des données du département vers l'OpenData et le suivi des réponses à la demande. Les cinq autres sont ponctuelles.

La première mission que j'ai eue à mon arrivé dans le département est l'écriture des fiches 36 et 37 pour la brochure EESRI 2021. Elle a duré un mois, elle a commencé en mi - octobre et a fini en mi-novembre. Cette mission a pour but de produire un rapport annuel sur l'Etat de l'enseignement supérieur et de la recherche et innovation en France. Je me suis occupé des parties sur la parité et sur les moyens humains dans la recherche en France. Pour cette mission j'ai travaillé principalement sur Excel et j'ai collaboré avec tout le département que ce soit du côté entreprise ou administration. Les difficultés sont liées à l'environnement, il m'a fallu un temps d'adaptation pour savoir où aller chercher les informations. Les deux fiches qui me sont confiées ont été réalisées dans le délai imparti.

Les deux missions de relecture sont dans la continuité de cette mission. Elles se sont déroulées au mois d'avril pour la version Web et au mois de juin pour la version papier. Toutes les deux ont une durée d'une semaine. Elles ont pour but de corriger les éventuelles erreurs et de définir la version finale avant de les diffuser au grand public. Je n'ai pas rencontré de difficulté particulière, toutes mes remarques ont été prises en compte dans la version finale. La version Web est déjà présente sur le site du Ministère depuis le mois d'avril. Les deux missions ont été réalisées dans le délais imparti.

La seconde mission que j'ai eu était l'intégration des données du département recherche vers l'OpenData. C'est un travail que j'ai effectué tout au long de l'année. Cette mission a pour but de transformer les fichiers de données obsolètes sur la recherche et développement dans les entreprises, dans les collectivités territoriales vers un format plus exploitable afin de les intégrer sur OpenData, une plateforme ouverte de données du Ministère. J'ai utilisé principalement Excel comme outil de production, j'ai collaboré avec le responsable des enquêtes côté entreprise Thomas Balcone et aussi avec le département des outils d'aide à la décision du SIES. La principale difficulté de la mission est de savoir comment transformer les données et sous quel format. On a eu plusieurs communications par mail avec le département des outils d'aides à la décision avant de définir le format adéquat.

Une autre difficulté est la quantité de données à transformer. Pour chaque fichier, nous avons entre 500 et 3 000 lignes de données, pour chaque fichier je dois vérifier s'il y a des erreurs. Si c'est le cas je dois recalculer les agrégats (par exemple la DIRDE, DERDE, ...) et de vérifier avec les données qui sont déjà publiées ou qui sont présentent sur la base de données du département. J'ai essayé d'écrire un script en Python pour automatiser la tâche, mais comme

chaque fichier est différent avec des codes, des noms, des tableaux différents et qu'il faut parfois recalculer des données, des agrégats, le résultat n'est pas du tout satisfaisant. Donc j'ai tout refait à la main.

Au final, j'ai transformé 44 fichiers, dont plusieurs sont intégrés sur la plateforme OpenData et d'autres sont en attente de validation.

La troisième mission qui m'a été confiée c'est la mise en place du suivi des réponses à la demande. Elle a commencé en janvier, le département reçoit tous les ans des demandes extérieures, que ce soit sur les enquêtes ou sur les données. Nous avons mis en place ce suivi car avant les demandes sont directement envoyés aux chargés d'études et on ne sait pas qui a répondu à telle demande et on ne sait pas quand la réponse a été envoyée. Après l'organisation de ce suivi, toutes les demandes sont centralisées via la boîte fonctionnelle du SIES. Nous organisons toutes les semaines une réunion pour connaître l'état d'avancement des demandes et de classer les demandes récentes.

Les deux missions de réponse à la demande sont incluses dans cette mission, elles ont eu lieu au mois de février pour la demande des données de R&D sur les Dom Tom, et au mois d'avril pour la demande concernant les données chercheuses dans la région des Bourgogne-Franche-Comté. Je n'ai pas éprouvé de difficulté particulière pour la réalisation de ces missions. Toutes les données sont trouvables dans notre base de données, sur le site Repères et Eurostat. J'ai fait la synthèse des éléments qui sont à ma disposition et j'ai répondu.

Conclusion

Durant cette année d'apprentissage au sein du département des études statistiques de la recherche, en participant à la réalisation nombreuses missions importantes, m'a permis d'acquérir une méthode de travail rigoureuse que ce soit dans le monde professionnel ou dans mes études.

Je suis conscient de la chance que j'ai de pouvoir effectuer cette année d'alternance au sein d'une structure qui m'offre la possibilité de voir comment les statistiques et les données sont utilisées pour évaluer la performance et l'investissement d'un pays sur la recherche, le développement et l'innovation.

Lors de la mission sur la brochure EESRI, j'étais un peu perdu au début avec les données à trouver car je ne connaissais pas très bien le sujet et l'environnement dans lequel je travaille. Mais ensuite j'ai su m'adapter en consultant les travaux faits par les autres personnes avant moi et avec l'aide des collègues, notamment avec Stéphane MONTENACHE mon collègue de bureau. J'ai acquis des compétences relationnelles à travers les échanges avec les autres membres de l'équipe. Les relectures des fiches m'ont appris à faire attention aux détails, la moindre confusion peut mener à un mal interprétation des données.

Pour la mise en forme des données vers OpenData, cette mission m'a permis de voir quel est le travail d'un statisticien. Je suis également devenu beaucoup plus rigoureux lorsque je traite les données, car elles seront utilisées par un grand nombre de personnes. Leurs travaux seront basés sur les données statistiques que je suis en train de traiter.

Cependant, je souligne un point sur le rythme d'alternance. L'alternance dans la semaine n'est pas une chose facile, il y a certaines réunions de départements auxquelles je n'ai pas pu assister car elles se déroulent durant ma période de cours. De plus alterner entre l'école et l'entreprise dans la même semaine pendant les périodes d'examens est un peu difficile à tenir, car les projets et les révisions demandent beaucoup d'énergie. Je pense que la situation sanitaire exceptionnelle que nous vivons aujourd'hui a également un grand impact.

Mais, ces contraintes m'ont permis de m'organiser plus efficacement et d'avoir une meilleure gestion de mon travail.

La période en présentielle était courte au début à cause de la crise sanitaire, mais depuis le lancement de la vaccination, nous sommes de plus en plus nombreux à y retourner. Travailler au bureau était bien plus confortable, de plus j'ai fait de nouvelles connaissances avec les nouveaux collègues apprentie, vacataire et stagiaire.

Cette année d'apprentissage m'a conforté dans le choix de mon orientation après le DUT. Je me suis orienté vers un poste et un domaine qui me permet de faire du traitement statistique des données et en même temps me permet d'approfondir mes compétences en informatique. Avec les données sur la R&D que j'ai traitée au cours de cette année et avec la participation au Funathon, je sais maintenant un peu plus sur la situation de la recherche en France notamment en informatique, en IA et en science des données. Je m'orienterai peut-être à la fin de mon cursus d'ingénieur vers une thèse dans le domaine de l'IA, notamment sur la reconnaissance d'image et de texte.

Annexe 1

En 2018, les collectivités ont dépensé 1,51 Md€ en faveur de la recherche et de l'enseignement supérieur

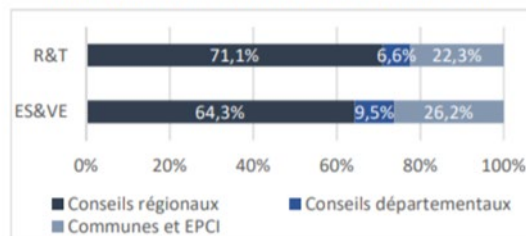
En 2018, les collectivités territoriales déclarent avoir dépensé 1,51 milliard d'euros (Md€) au profit de l'enseignement supérieur et de la recherche (ESR), soit un niveau similaire à celui de 2017 (1,52 Md€). Les dépenses en faveur de l'enseignement supérieur et de la vie étudiante (690 M€) sont, pour près de la moitié, constituées de dépenses immobilières, tandis que les aides aux entreprises innovantes et au transfert représentent 36 % de celles couvrant la recherche et le transfert de technologie (820 M€). En 2019, le budget prévisionnel s'établissait à 1,7 M€.

Une hausse des dépenses d'ESR des collectivités territoriales en 2019

En moyenne, sur la période 2017 – 2019, les collectivités territoriales auraient consacré près de 1,6 Md€ à l'enseignement supérieur et à la recherche, selon les résultats de la dernière enquête réalisée. Stables en 2018 à 1,51 Md€ (1,52 Md€ en 2017), ces dépenses devraient avoir progressé en 2019 (1,7 Md€, données provisoires). Les dépenses concernant l'enseignement supérieur et la vie étudiante (ES&VE) s'accroissent entre 2017 et 2018, passant de 580 M€ à 690 M€, tandis que pour la recherche et le transfert de technologie (R&T), les dépenses se replient : 820 M€ en 2018 après 940 M€ en 2017. L'année 2019 devrait présenter un accroissement dans les deux domaines (800 M€ pour l'ES&VE et 910 M€ pour la R&T).

Dépenses d'ES&VE et de R&T des collectivités territoriales au cours de la période 2017 – 2019

Répartition, entre les collectivités territoriales, du financement moyen sur les trois exercices de l'ES&VE et de la R&T



Données semi-définitives (moyenne 2017-2019)

Source : MESRI-SIES.

Deux territoires régionaux concentrent plus de 30 % du financement de l'ESR par les collectivités territoriales

Les collectivités territoriales d'Île-de-France financent l'ESR à hauteur de 325 M€, en moyenne sur la période, ce qui

Depuis près de 30 ans, le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation organise annuellement, avec la contribution de plus de mille porteurs de projet, plus de 3 000 événements au cours de sa traditionnelle Fête de la science. Ces événements, coordonnés aux échelles nationales, régionales et départementales, sont mis en œuvre en partenariat avec 7 000 structures scientifiques, culturelles et institutionnelles (organismes de recherche, établissements d'enseignement de tout niveau, collectivités territoriales, entreprises, musées, bibliothèques, médiathèques, associations...) et l'aide de 33 000 acteurs. Ils permettent la rencontre

Bilan de la fête de la science 2019

La crise sanitaire vécue en 2020 a rappelé à chacun le rôle majeur de la recherche vis-à-vis de l'économie et de la société. Un travail de pédagogie est ainsi nécessaire pour faire percevoir tout l'apport de la science sous ses différentes facettes, et la manière dont elle se développe. Aussi, tout au long de l'année, chercheurs, enseignants, mais également responsables d'entreprises ou d'associations et bénévoles s'engagent pour partager les sciences avec le plus

une place centrale en matière de culture scientifique, par le nombre, la richesse et la variété des événements organisés simultanément sur tout le territoire français. La Fête de la science a plusieurs objectifs : éclairer les citoyens sur les avancées de la recherche et ses enjeux, leur donner les moyens de renforcer leur curiosité et leur esprit critique grâce à la démarche scientifique et susciter des vocations, notamment chez les jeunes, quels que soit

Annexe 3



[Accueil](#) | [Qui sommes-nous ?](#) | [Chiffres](#) | [Publications](#) | [Contrats](#) | [Formulaires](#) | [REPERES](#) |
[Liste de diffusion](#) | [Chercher](#) | [FAQ](#) | [Sigles](#) | [Plan du site](#) | [Contacts](#) | [Nouveau](#)

À la une

Les prévisions 2018

» [Note flash Enseignement supérieur, Recherche & Innovation du SIES \(n° 20.05 avril\)](#)

[Les dépenses de R&D des entreprises en 2018 en France \(données provisoires\)](#)

Brochure - [Vers l'égalité Femmes-Hommes ? Chiffres clés \(8 mars 2020\)](#)

Quelles sont les différences observées entre les jeunes femmes et les jeunes hommes se dans leurs choix de formations supérieures? En quoi et à quel moment les carrières académiques des femmes et des hommes divergent elles ? Dans quelle mesure les femmes sont-elles encore peu présentes dans les organes de gouvernance ? La situation en France se distingue-t-elle des autres pays ? Quel est le chemin qui reste à parcourir sur le chemin de l'égalité dans l'enseignement supérieur, la recherche et l'innovation ? Cette quatrième édition permet d'offrir à l'ensemble des acteurs et au public des données détaillées et objectivées des différences observées dans le champ de l'enseignement supérieure, de la recherche et de l'innovation, afin que chacun puisse s'emparer de ces données, les diffuser et œuvrer à des actions concrètes en faveur de la parité.

Annexe 4



REPERES R&D

	STATISTIQUES	PUBLICATIONS	QUESTIONNAIRES D'ENQUÊTES	DÉFINITIONS, MÉTHODES ET QUALITÉ
---	--------------	--------------	---------------------------	----------------------------------

ACTUALITÉS

Accueil du site Repères R&D.

Une description de cette page : Lorem ipsum dolor sit amet, consectetur adipiscing elit. Proin porttitor lorem eu justo mattis, feugiat lacinia lacus porta. Nulla egestas nunc nec sapien vestibulum pharetra. Cras egestas, tortor ac venenatis ornare, diam mauris ullamcorper dui, a mattis dolor urna quis sem. Etiam et dui nec felis accumsan suscipit. Phasellus sed mattis orci. Sed sagittis molestie posuere. Maecenas eleifend porta ipsum, vitae finibus neque accumsan at. Aliquam quis pharetra nibh, sed auctor arcu. Nulla non libero nec ante semper sagittis. Sed hendrerit odio in fringilla tristique. Sed imperdiet faucibus enim, eget tristique eros tempus eget.