

Lyon, 7 mars 2024

— université
— lumière
— LYON 2

UFR **anthropologie**
sociologie
science politique

eric

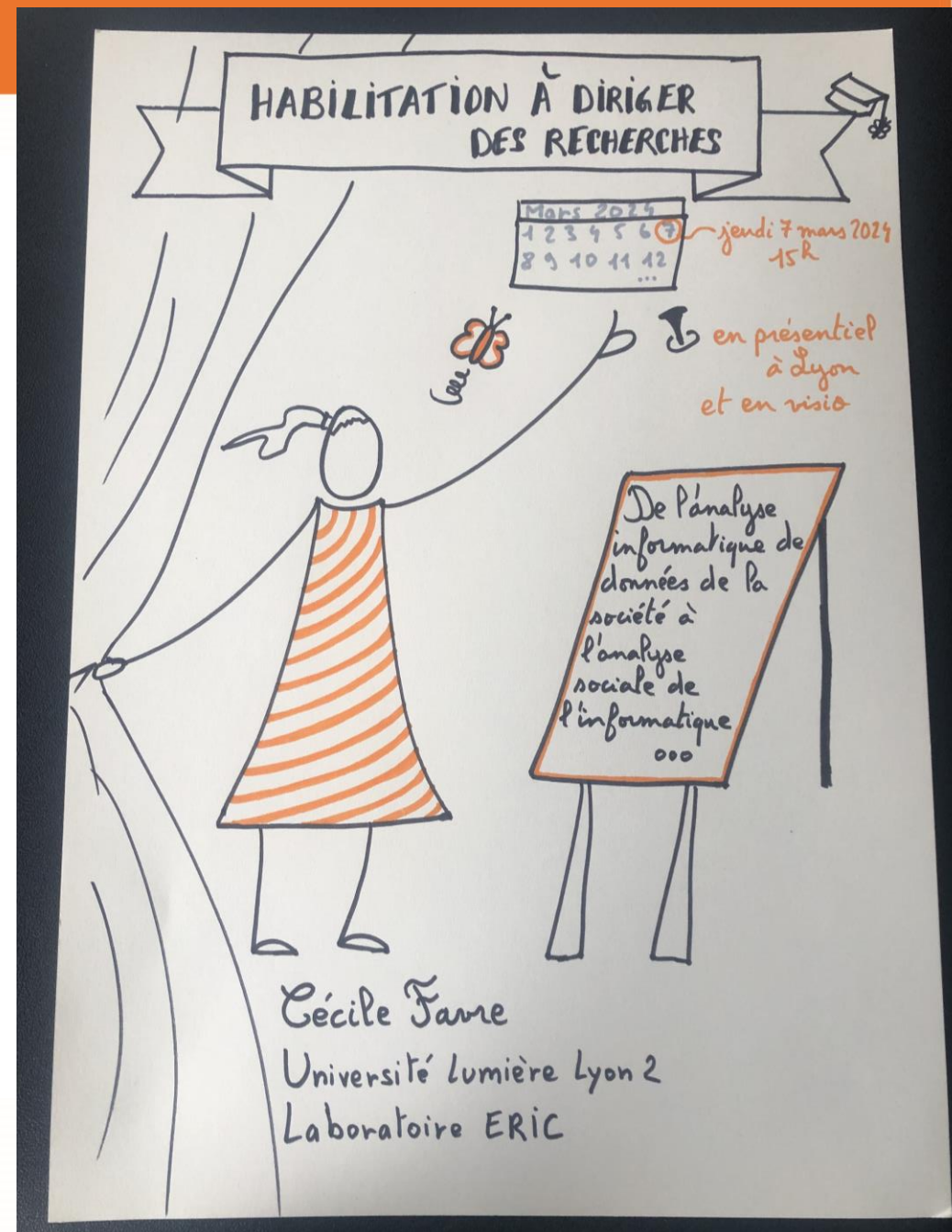
Centre
Max Weber

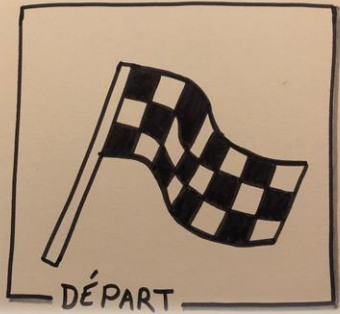
UQAM IREF
Institut de recherches
et d'études féministes

De l'analyse informatique de données de la société à l'analyse sociale de l'informatique.

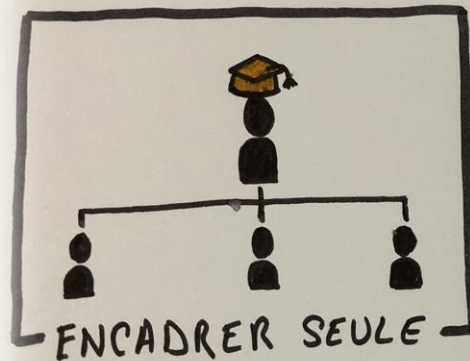
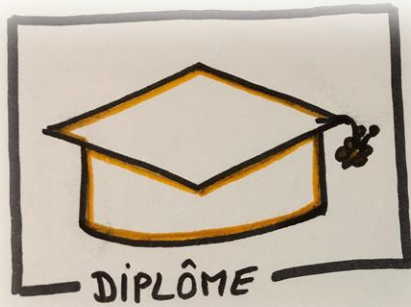
Un cheminement guidé par les études de genre
vers un décloisonnement disciplinaire
et une posture réflexive.

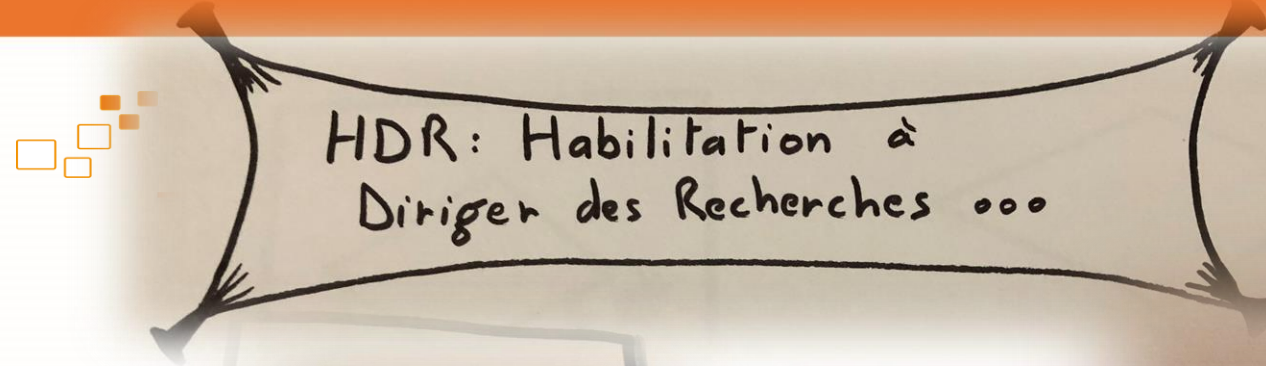
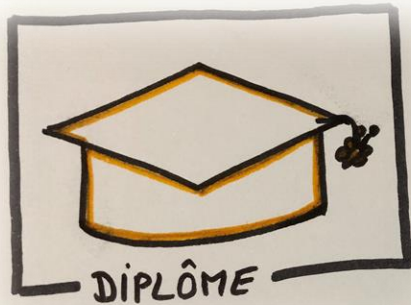
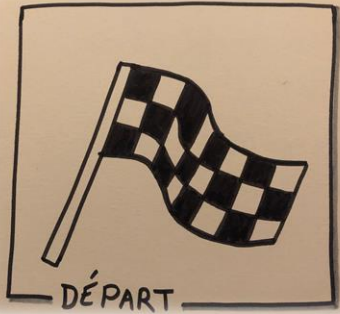
Soutenance pour l'obtention de
l'Habilitation à Diriger des Recherches
Garant : Jérôme Darmont





HDR: Habilitation à
Diriger des Recherches ...







Le clin d'œil du lieu



■ Amphi BENVENISTE

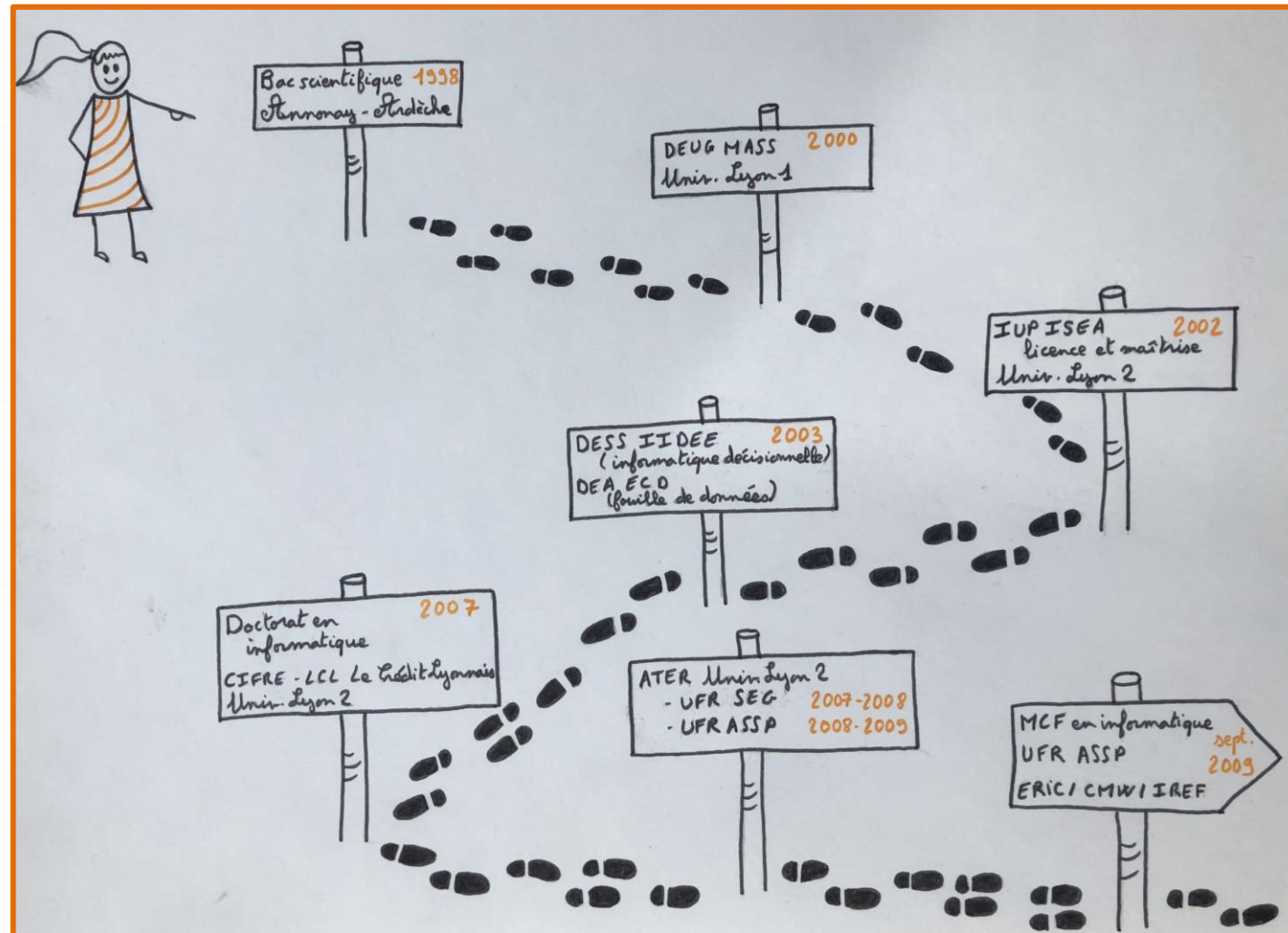
- Émile Benveniste
(Alep : 1902 – Versailles : 1976)
- Linguiste français :
né Ezra Benveniste syrien,
naturalisé français en 1924
- Travaux reconnus sur une
méthode d'analyse du langage
mettant en avant **la dimension
sociale** de celui-ci, amenant des
résultats qui intéressent
l'histoire et l'anthropologie





Un parcours vers l'informatique et les sciences sociales

Reparcourons le chemin !

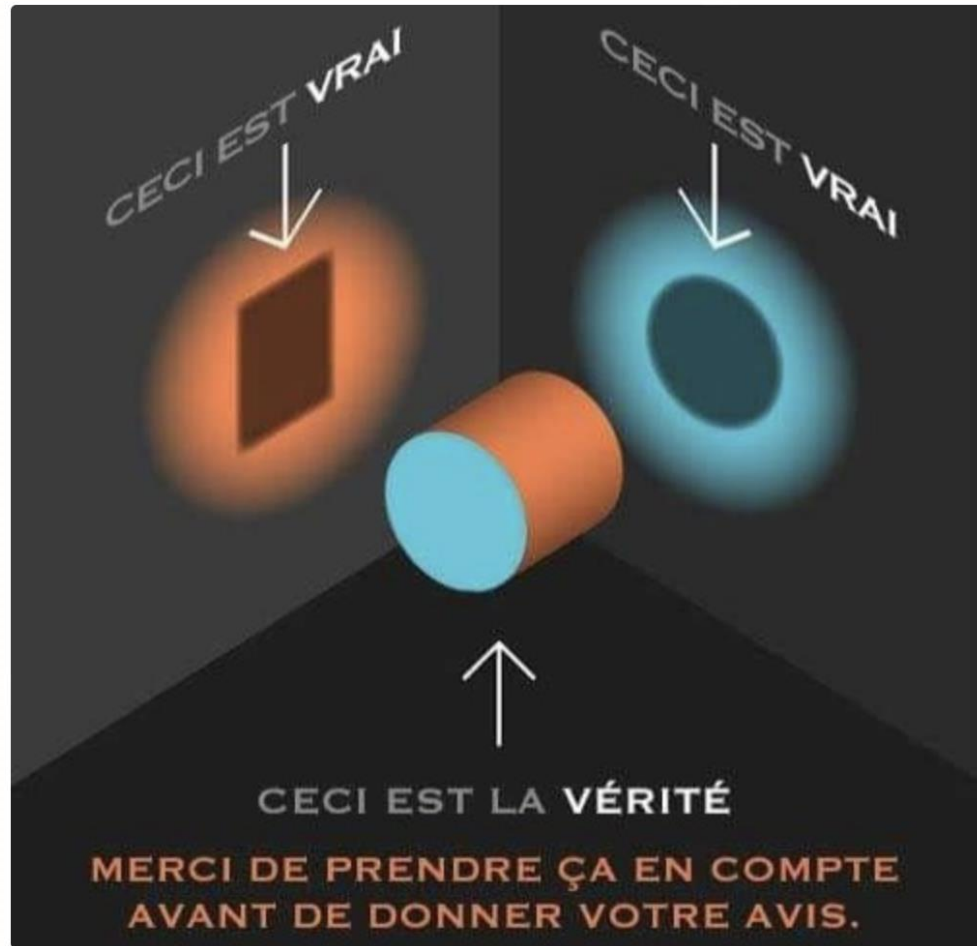


Au sein de l'UFR ASSP

(Anthropologie, Sociologie, Science Politique) :

- Implication forte en Licence **MIASHS**
- Responsabilités croissantes dans la Mention de Master **Études sur le genre** de Lyon
- Multiplier les points de vue et changer de regard...

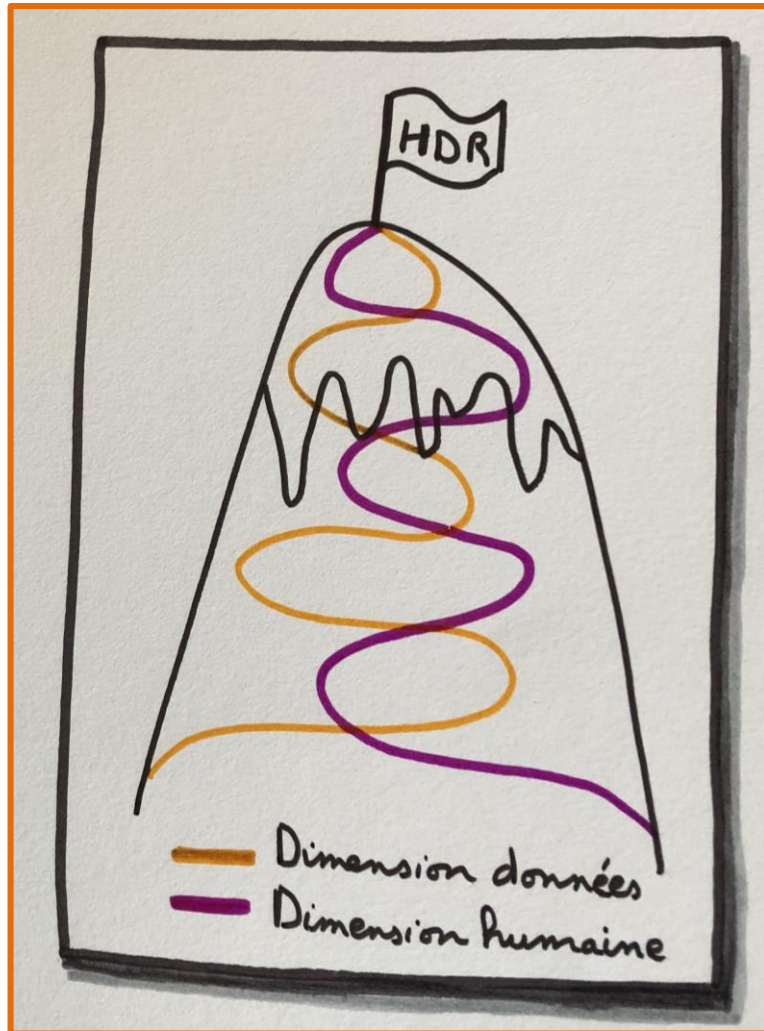
Tout est question de points de vue !



■ Illustration par le concret :

- Processus d'analyse de données
- Posture de recherche
- Invitation à changer notre regard
 - Autre manière d'aborder la restitution des contributions et des perspectives par rapport au manuscrit

Tout est question de points de vue !



■ Illustration par le concret :

- Processus d'analyse de données
 - Posture de recherche
 - Invitation à changer notre regard
- Autre manière d'aborder la restitution des contributions et des perspectives par rapport au manuscrit

Plan de vol



Survol des travaux réalisés



Planter le décor : le contexte d'encadrement



Pour des utilisateur/trices métier



Pour des chercheuses et chercheurs



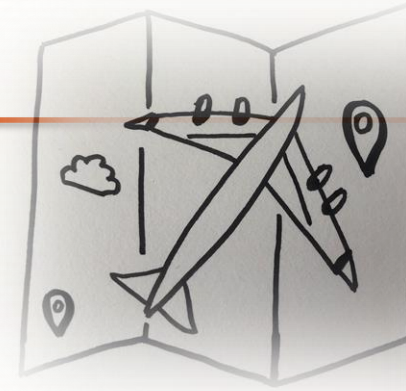
Regarder en arrière : ce que j'ai appris de mon expérience

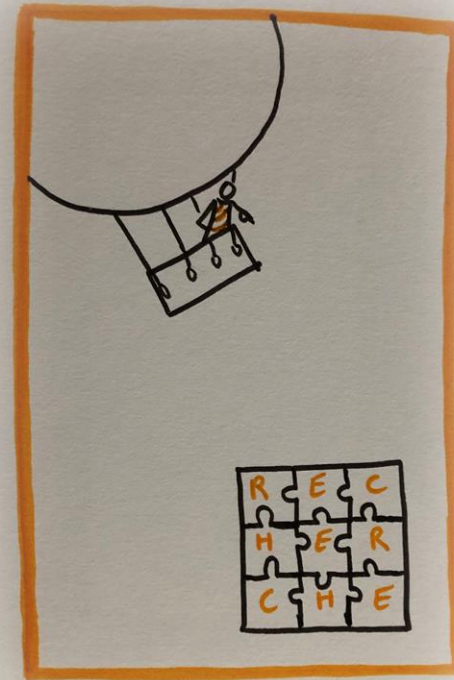


Regarder devant : programme de recherche

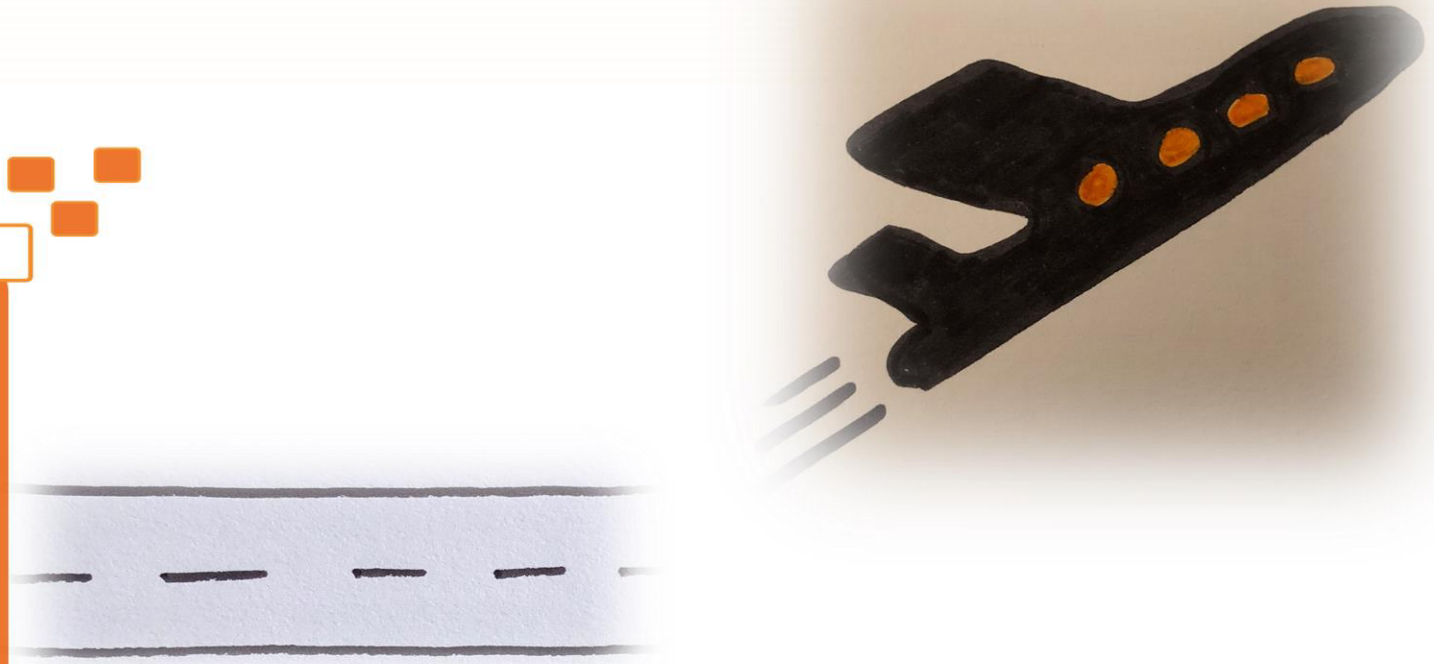


Pour conclure





Survol des travaux réalisés

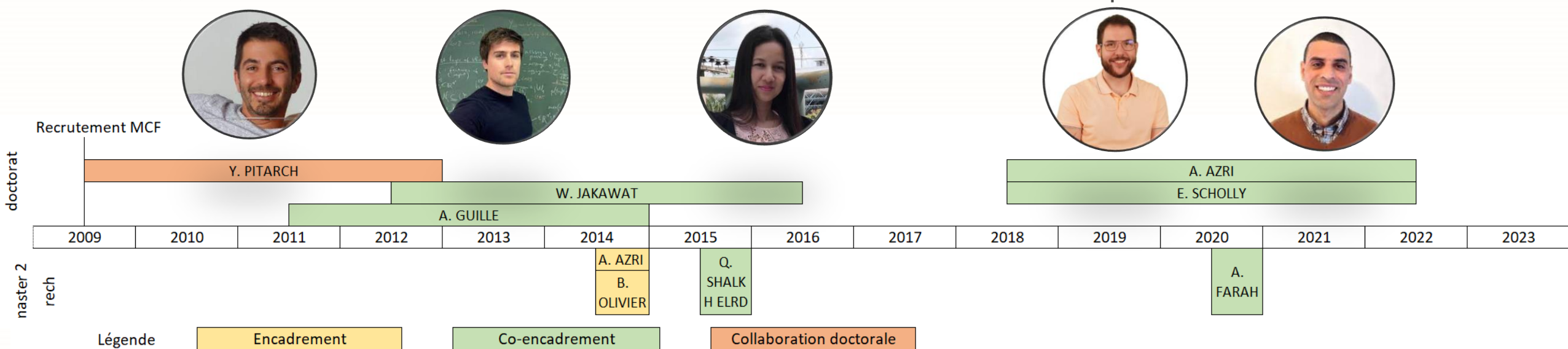


Survol des travaux réalisés

Planter le décor : le contexte d'encadrement

Planter le décor : le contexte d'encadrement

(Se) construire dans le collectif : des co-encadrements doctoraux et bien plus !



- Yoann Pitarch : Résumé de Flots de Données : motifs, Cubes et Hiérarchies
- Etienne Scholly : De la modélisation des métadonnées à la conception d'un lac de données : Application à l'habitat social
- Adrien Guille : Diffusion de l'information dans les médias sociaux : modélisation et analyse
- Abderrazek Azri : Approches multimodales d'apprentissage automatique pour la détection des rumeurs dans les microblogs
- Wararat Jakawat : Graphs enriched by Cubes (GreC): a new approach for OLAP on information networks



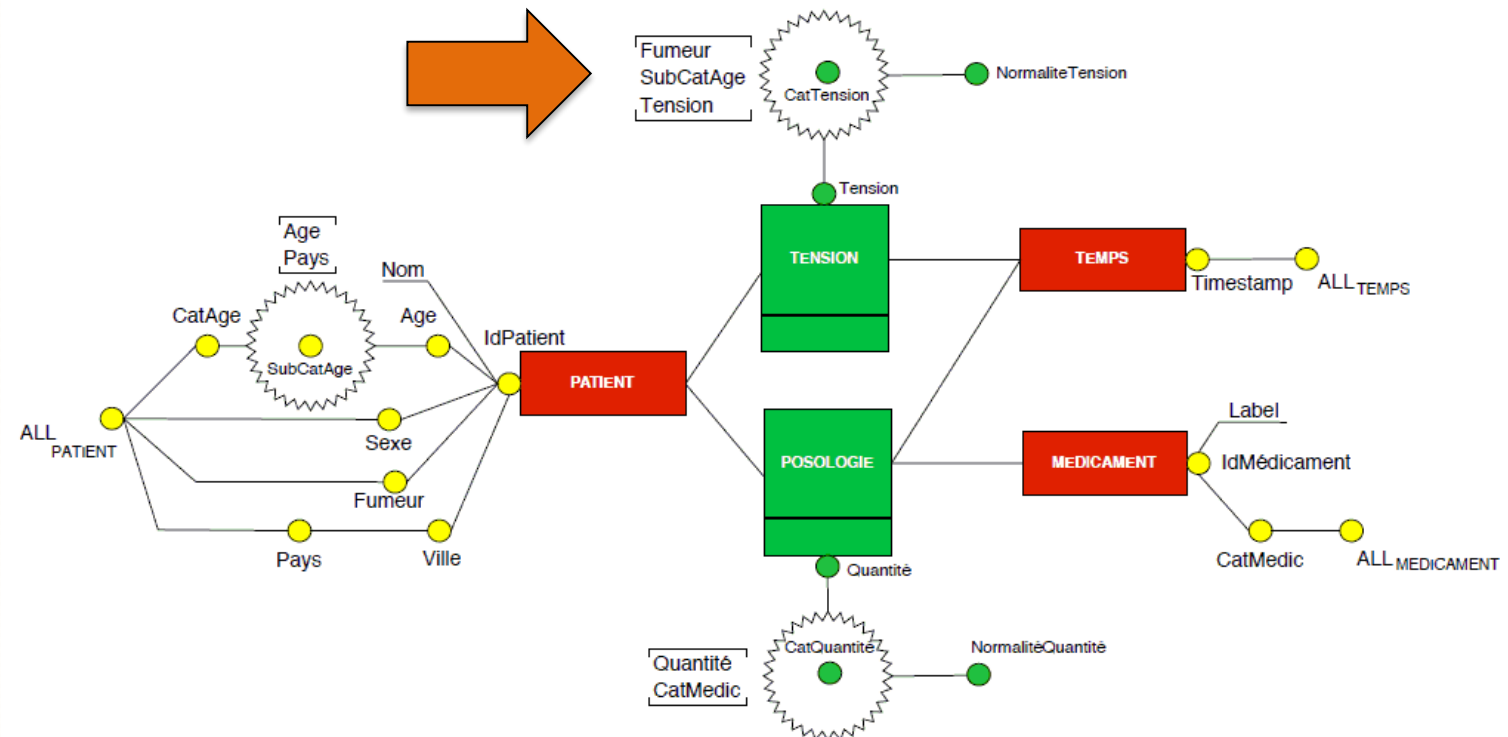
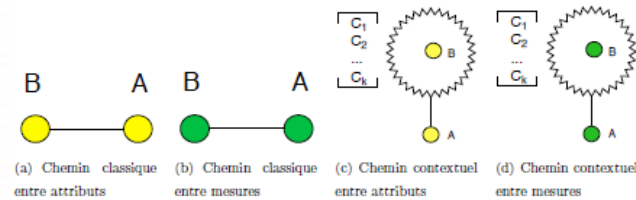
Survol des travaux réalisés

Pour des utilisateur/trices métier



Modélisation multidimensionnelle et connaissances métiers

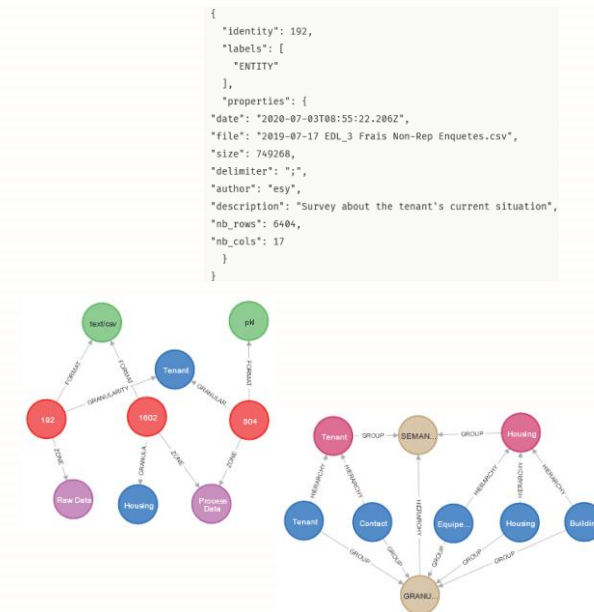
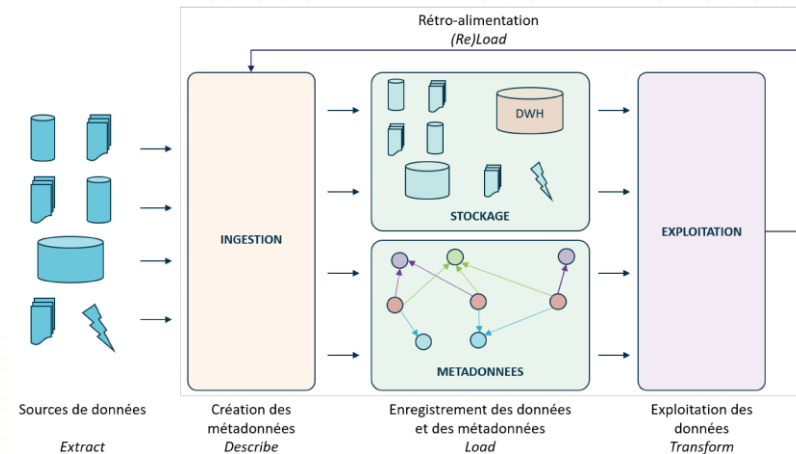
- Expression des connaissances des utilisateur/trices métier pour l'analyse : contexte du médical
- Travail de thèse de Yoann Pitarch encadré par Anne Laurent et Pascal Poncelet, soutenue en 2011
- Enrichissement du modèle multidimensionnel par des « satellites » pour davantage de flexibilité



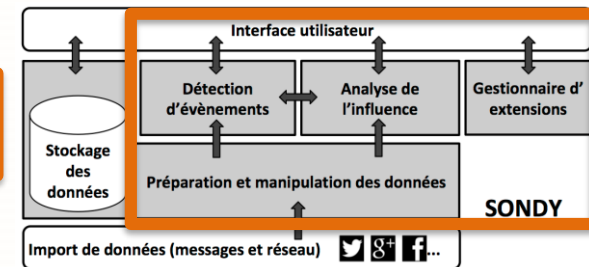
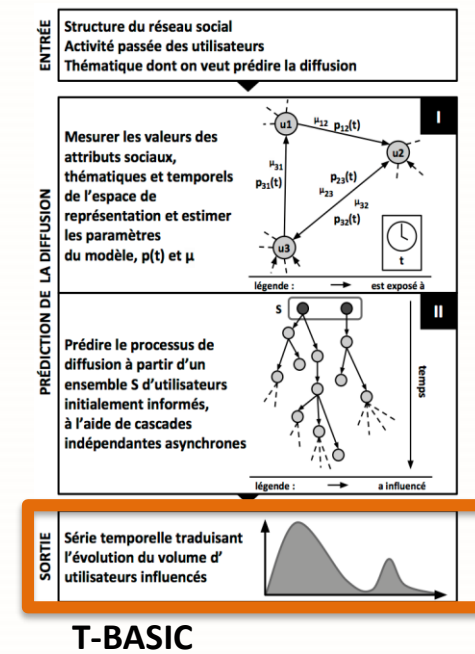
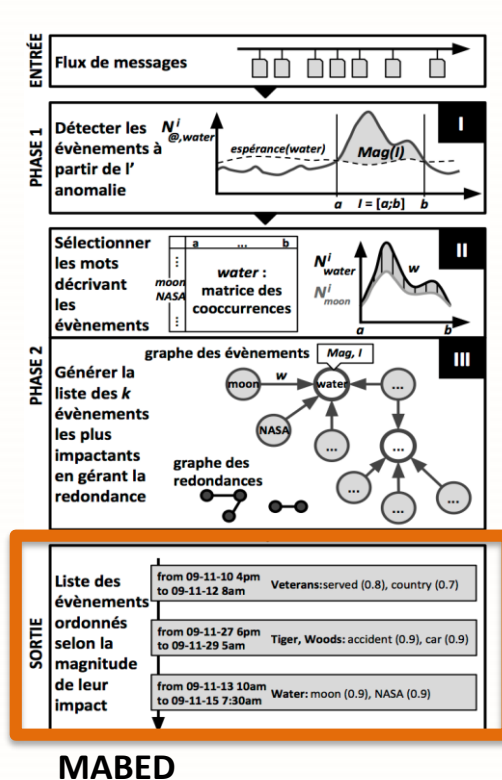
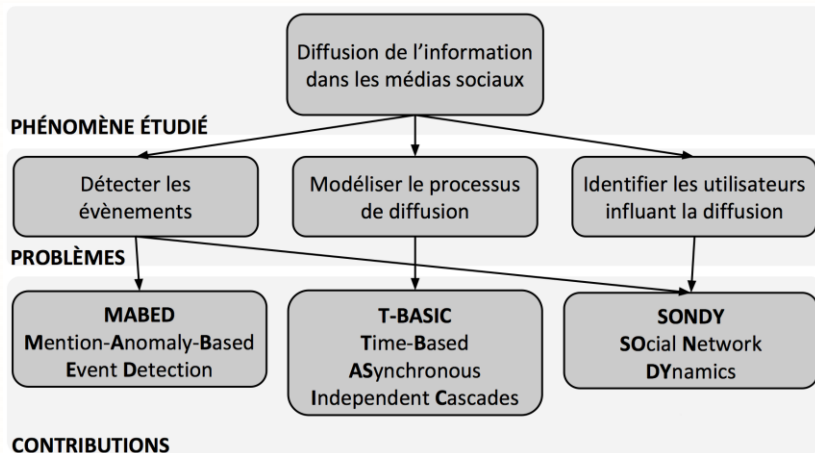
Lacs de données et leurs métadonnées

- **Répondre aux besoins d'analyse des utilisateur/trices métier : contexte de l'habitat social**
- Thèse d'Etienne Scholly, co-encadrée avec Sabine Loudcher, soutenue en 2022
- Modélisation des métadonnées d'un lac de données : MEDAL et goldMEDAL
- Avec l'équipe des *datalakers* (représentation de contextes multiples : HyperThesau – archéologie / Bretez – multidisciplinarité sur données multimédia / COREL & AURA-PMI – sciences de gestion)

Critère / Système	Entrepôt de données	Lac de données
Type de données stockées	Données structurées (ou de nature homogène)	Données de tous types
Intégration des données	ETL (Extract - Transform - Load)	ELT (Extract - Load - Transform)
Définition du schéma	À l'insertion (schema-on-write)	À l'interrogation (schema-on-read)
Flexibilité	Rigide, peu évolutif	Flexible, mais besoin d'un système de métadonnées
Type d'analyses	Analyses prédéfinies	Tous types d'analyses

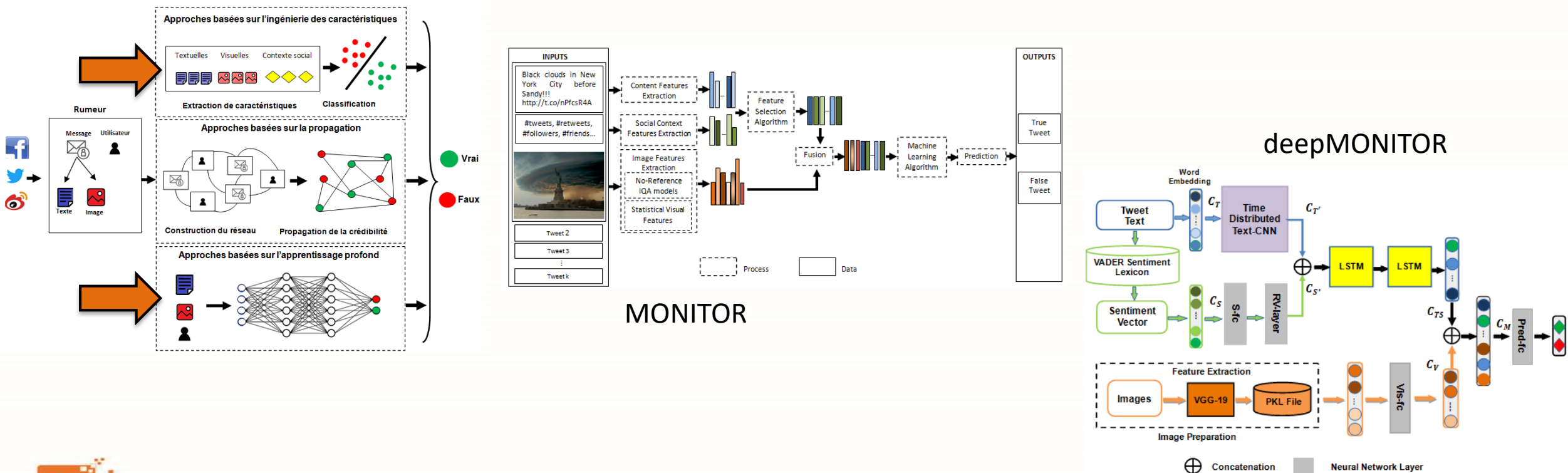


- **Aider le travail journalistique, les analystes de médias, ...**
- Thèse d'Adrien Guille co-encadrée avec Djamel Zighed, soutenue en 2014
- Contributions pour l'analyse de la diffusion de l'information dans les microblogs (Twitter)



Détection de rumeurs dans les médias

- Aider le travail journalistique, les analystes de médias sociaux, ...
- Thèse d'Abderrazek Azri co-encadrée avec Jérôme Darmont et Nouria Harbi, soutenue en 2022
- Approches multimodales de détection de rumeurs dans les microblogs (Twitter)





Survol des travaux réalisés

Pour des chercheuses et des chercheurs



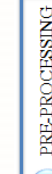
- 



C



...





Analyse de la communauté EGC

■ Poursuite de travaux en scientométrie : **amener une réflexivité sur sa propre discipline**

■ *Défis*

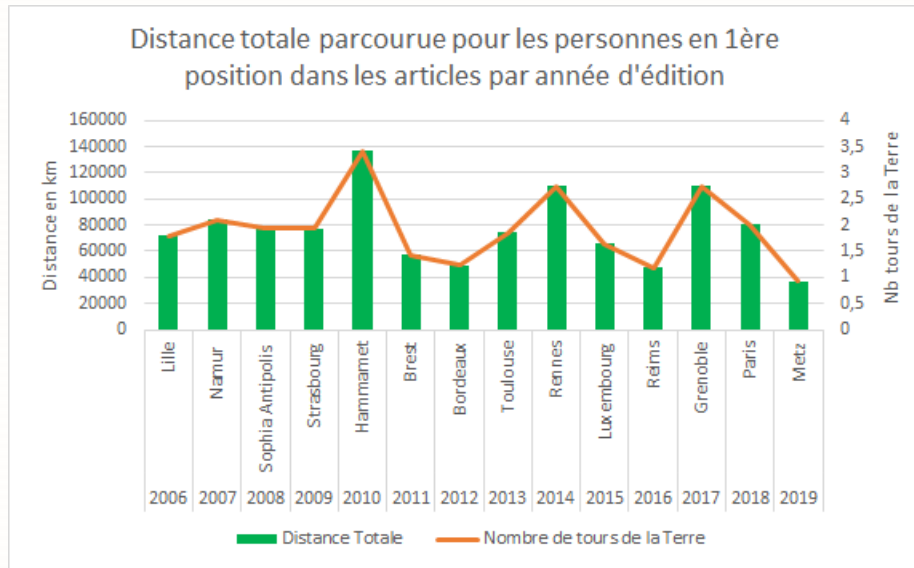
- Association EGC (Extraction et Gestion de Connaissances) avec conférence annuelle organisée en différents lieux
- Défi EGC 2016 « Communauté EGC : quelle histoire et quel avenir ? » + Défi EGC 2020 : « 20 ans d'histoire pour quel avenir ? » : mise à disposition des matériaux la concernant en vue de leur analyse plusieurs mois avant la conférence
 - Complétion de données et genrage de donnée
- Défi EGC 2020 : résultats dans l'article « **Regards d'actualité au prisme des enjeux sociétaux sur les données historisées d'EGC** », co-écrit avec Sébastien Valat



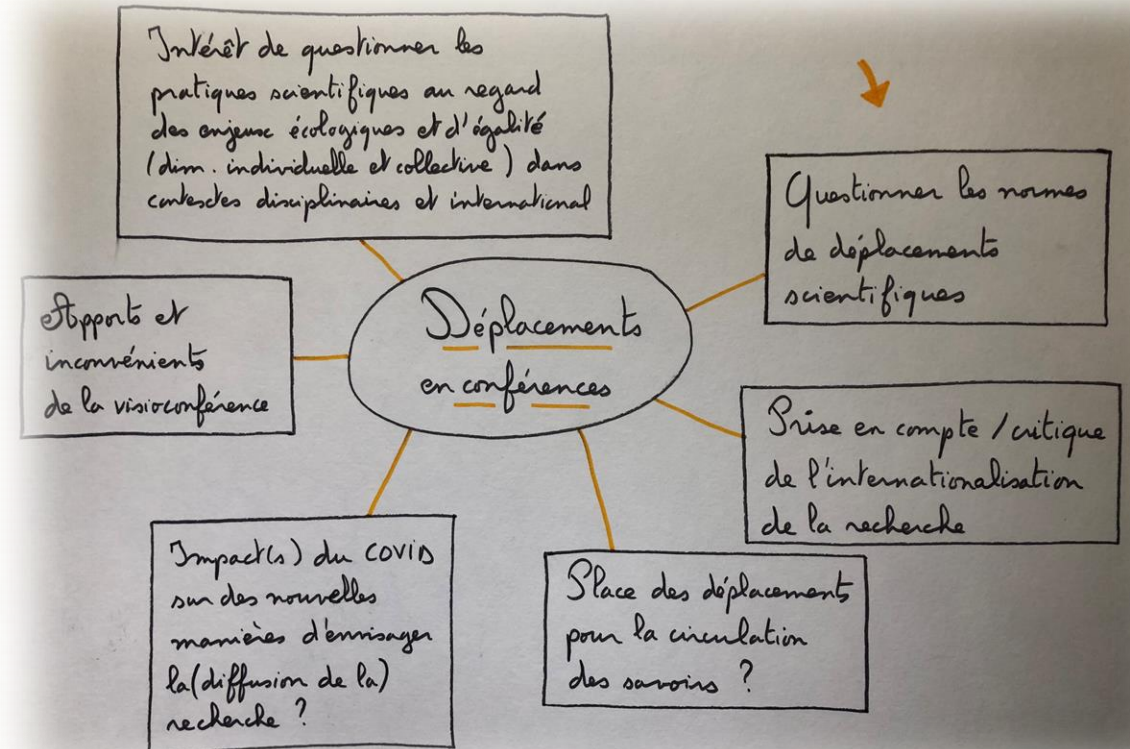


Analyse de la communauté EGC : dimension écologique

- Contexte : réchauffement climatique lié notamment aux émissions de carbone
- Analyse de la distance parcourue et de l'équivalent CO₂ émis (estimations basses) : personne en 1^{ère} position sur le papier avec automatisation de la géolocalisation à partir des PDF



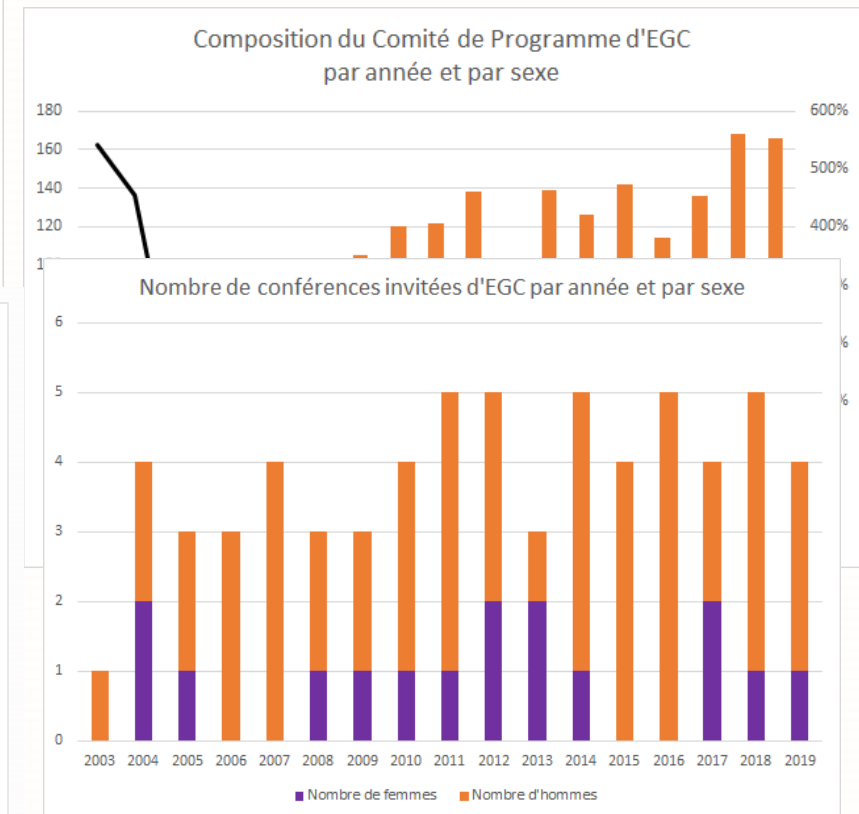
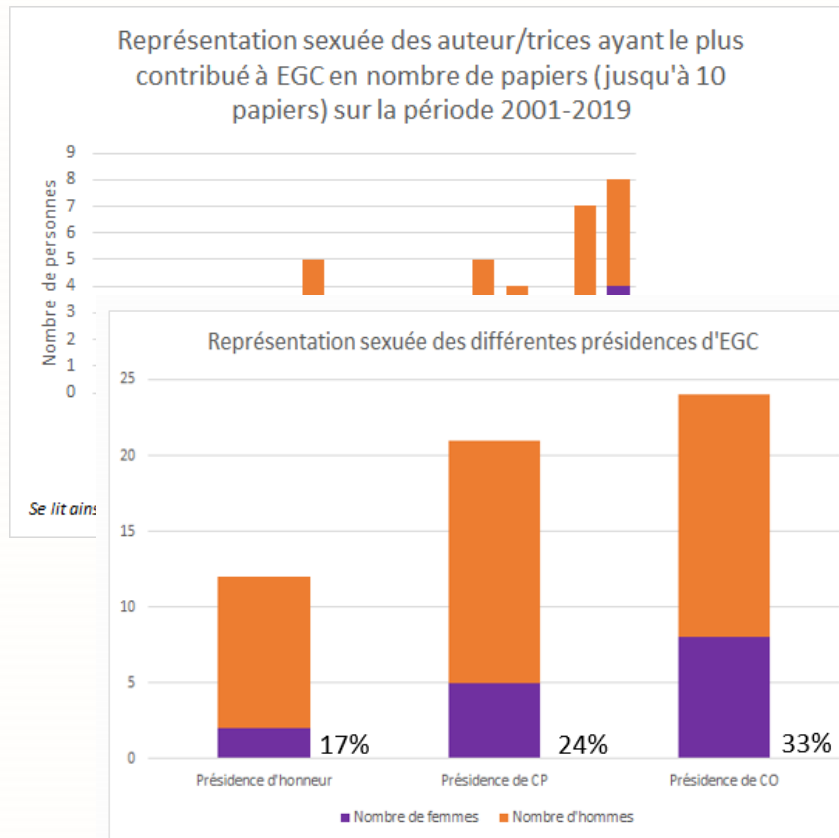
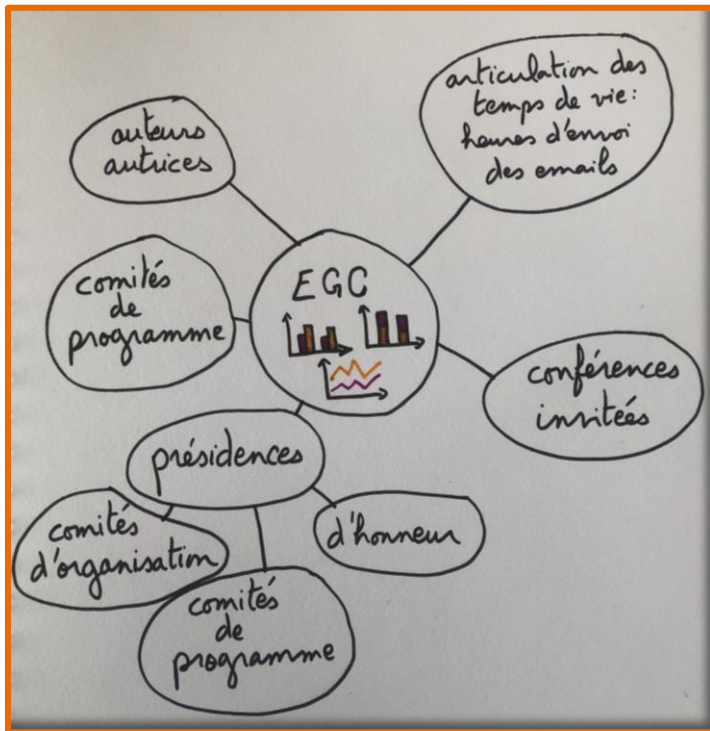
- Conférences : un enjeu de rencontres et d'échanges mais réflexion/discussion sur les déplacements scientifiques





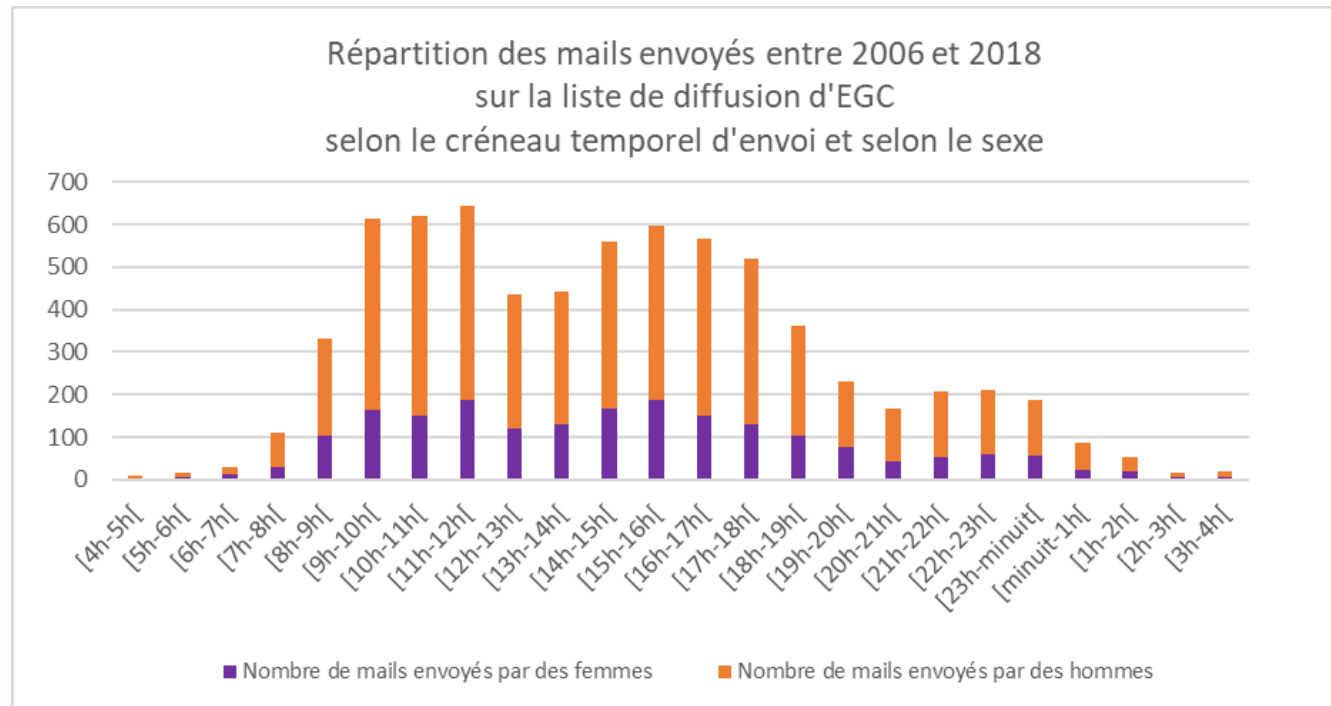
Analyse de la communauté EGC : place des femmes

- Contexte : sous-représentation des femmes ($\approx 25\%$)
- Enjeu : travailler de manière précise sur un jeu de données en abordant d'autres aspects que les publications



Analyse de la communauté EGC : place des femmes

- Contexte : travaux en SHS autour de l'épuisement au travail et de l'articulation des temps de vie, notamment chez les enseignant.es chercheuses et chercheurs / droit à la déconnexion
- Analyse des heures d'envoi des 7075 mails de 2006 à 2018 sur la liste de diffusion



1116 mails en dehors de la plage

8h-20h : **16% des mails**

Parmi **les mails hors période**

« classique » :

- **70% de 20h à minuit**
- **15% de minuit à 4h**
- **15% de 4h à 8h**

*Notification d'acceptation envoyée à **23h01**... !!!*

*Graphique réalisé un soir vers **23h00**... !!!*

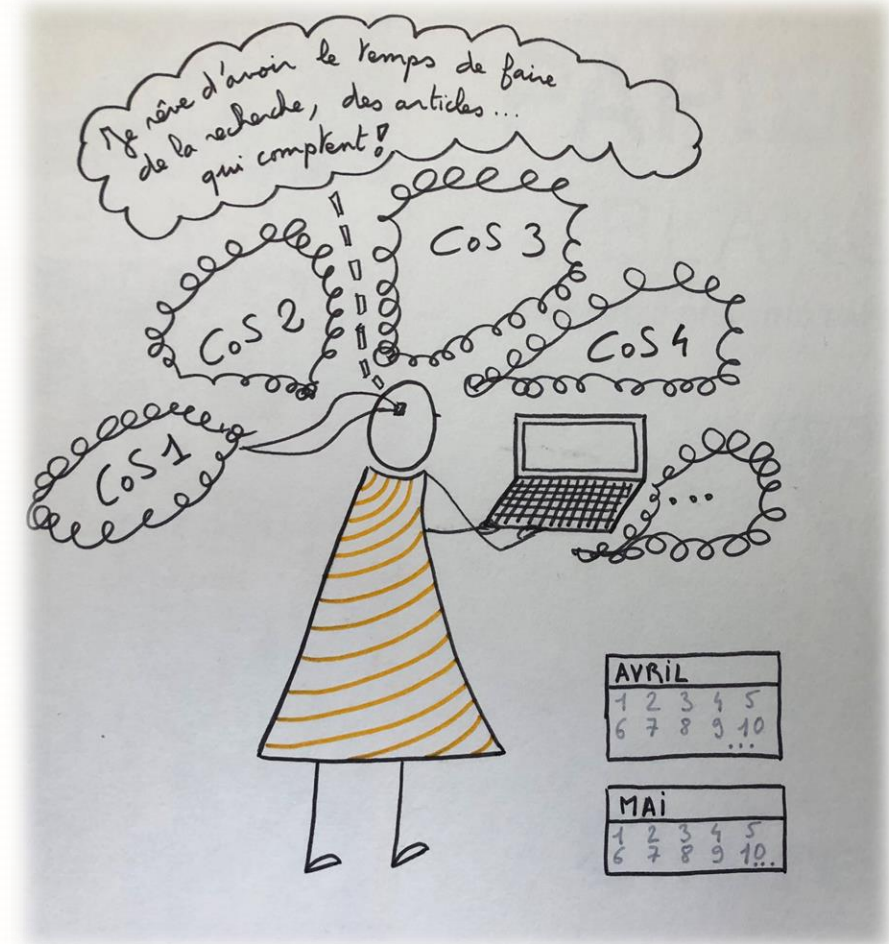


Analyse exploratoire sur les politiques de quotas

- Initiation à la démarche de recherche sociologique, en collaboration avec Laurence Tain
- Un « effet rebond » avec une politique d'égalité des quotas qui peut amener une sur-sollicitation des femmes dans les comités de sélection en informatique
- Sur-sollicitation des femmes ou plutôt **sur-sollicitation de certaines femmes !**

→ Proposition de mettre en place une ressource à construire façon « Les expertes »

<https://expertes.fr/>





Regarder en arrière :
ce que j'ai appris
de mon expérience



Place de la catégorisation de données

- Non neutralité du processus de catégorisation des données
- Impact en science des données
- Enjeu de visibilisation des catégories
 - Initiative de Statistique Canada à souligner : refonte des catégories en termes d'identité de genre et d'orientation sexuelle avec consultation publique
- Thématique non récente, traitée par de nombreuses disciplines, mais à réfléchir, requestionner

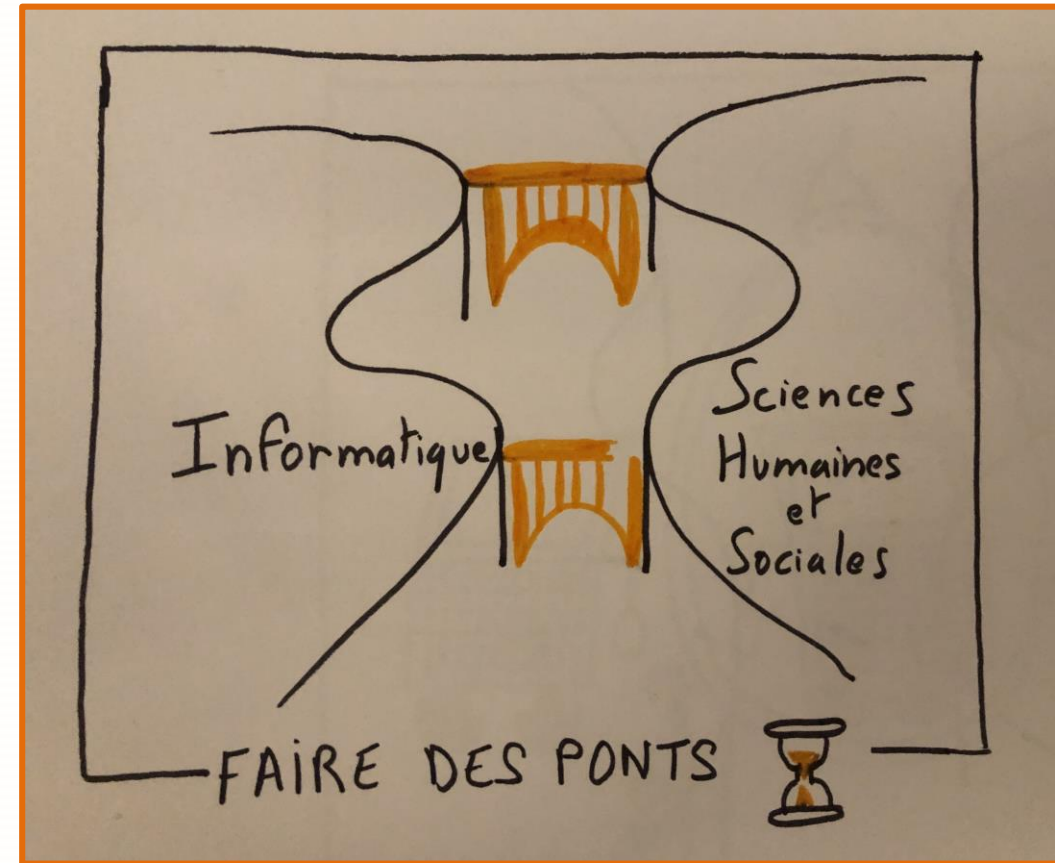


Cécile Favre, Marie Vialaret: **A Perspective on Data Categorization with Regard to Equity, Diversity, and Inclusion in Data Science**. Workshop DS4EIW (Data science for equality, inclusion and well-being challenges) in conjunction with IEEE Big Data 2023: 4156-4165.



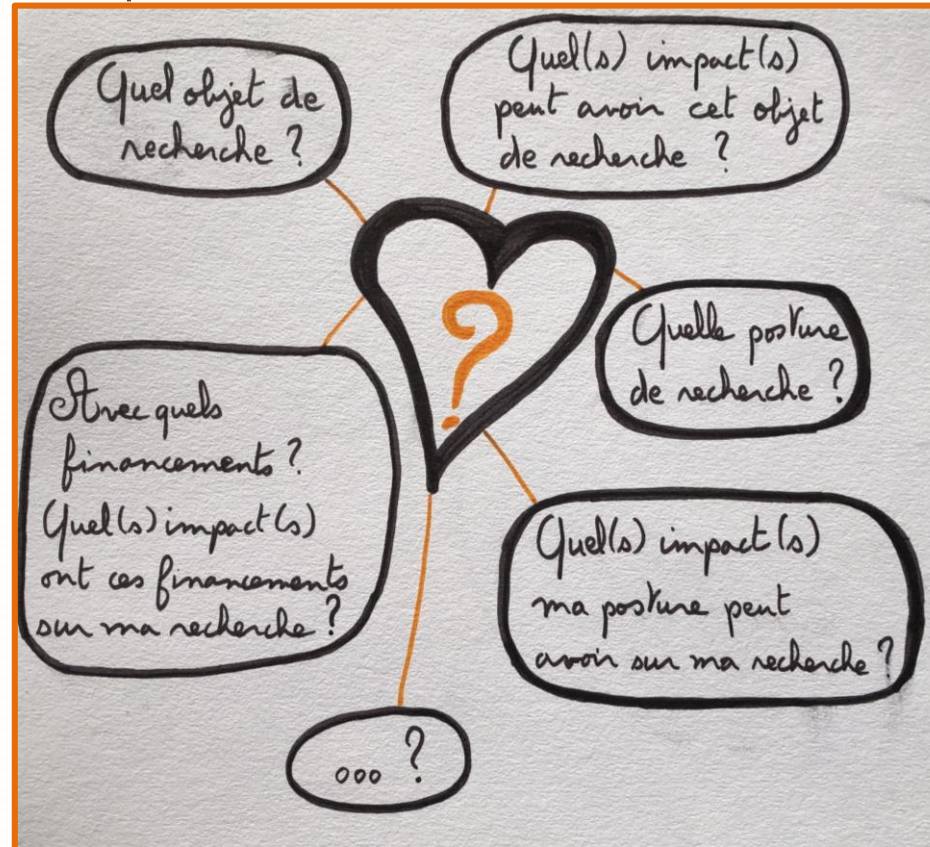
Nomadisme disciplinaire

- Ancrage en informatique
- Capacité à « arpenter » d'autres disciplines → Pas une errance
- Fil (parfois funambule) / ponts ...
- Conscientiser
 - Variation des normes disciplinaires (évaluation, ...)
 - Hiérarchies disciplinaires (notamment financement...)
- Conviction de la richesse du nomadisme :
→ apport d'une « largeur » en complément d'une « profondeur »
- Temporalité
 - Pour construire des ponts (solides)
 - De la recherche



Réflexivité

- Savoirs situés (Donna Haraway en 1988) : s'interroger sur la production des savoirs
- Remettre la question au cœur du processus de la recherche :
→ au-delà du « comment ? »



Epistémologie

- Enjeux des questionnements épistémologiques en informatique, pour une science jeune en quête de reconnaissance en tant que telle
- Apport d'autres disciplines (intérêt du nomadisme disciplinaire : dont la sociologie des sciences, l'histoire des sciences, la philosophie des sciences...)

Exemple en science de l'information et de la communication

Hubert Fondin, « La science de l'information : posture épistémologique, et spécificité disciplinaire », Revue « Documentaliste - Sciences de l'information », vol. 38, n° 2 (2001), p 112.

« en tant que chercheur en Science de l'information : quel phénomène voulons-nous comprendre ? Quelle est la place et quel est l'objet de cette science ? Quelle est la posture épistémologique des chercheurs qui, implicitement ou explicitement, revendiquent leur appartenance à cette science ? »

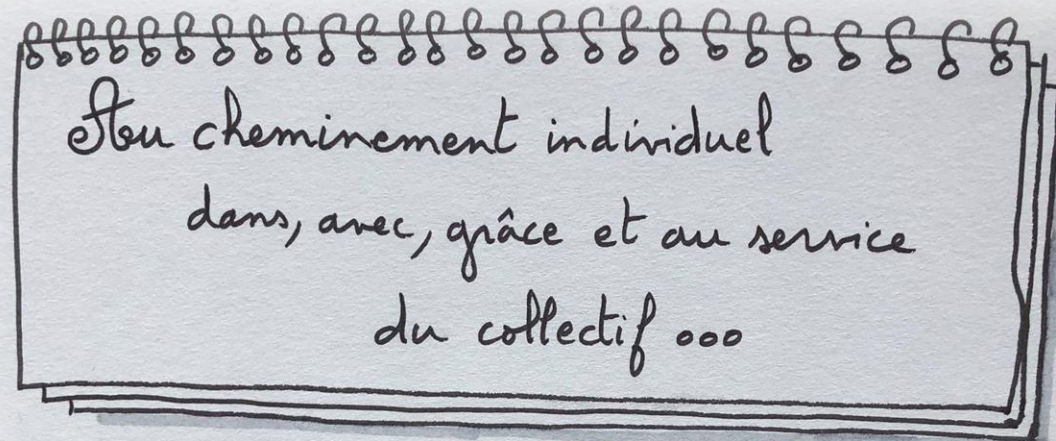


Retrouver du pouvoir d'agir

- Dimension systémique du fonctionnement de la recherche imparfait
- Certaines normes (évaluation) qui s'inscrivent sont subies (même si parfois dénoncées) : *publish or perish* !
- Questionner ces normes scientifiques (disciplinaires...)
- Retrouver du pouvoir d'agir sur ces normes scientifiques
- Exemple de l'internationalisation de la recherche et le regret que les communautés nationales ont des difficultés
 - Pourquoi pas une obligation d'une publication dans la communauté nationale pour obtenir la qualification aux fonctions de maître.sse de conférences ?

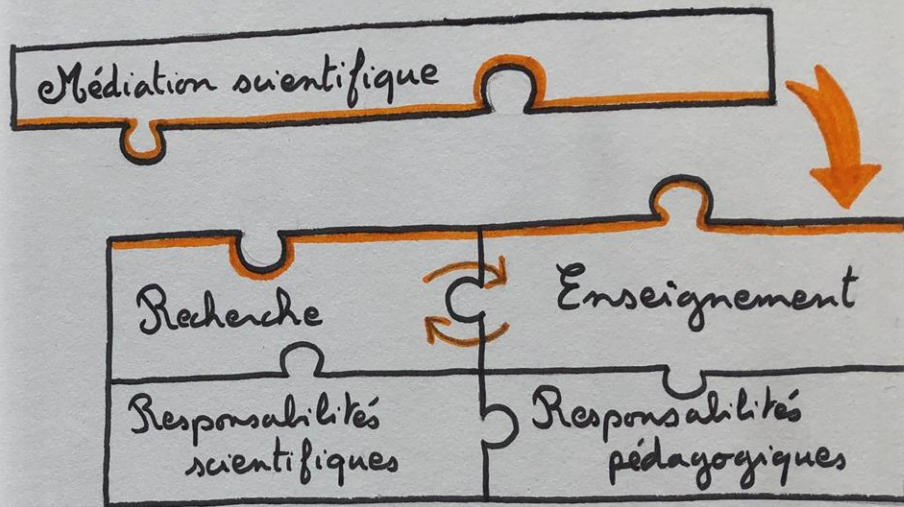
Science, sens, société et engagement

- Vouloir contribuer : quel sens ?
- Enjeu d'analyser les aspects de financement de la recherche qui ne sont pas neutres
- Avec quelles valeurs, quelle éthique, pour qui ?

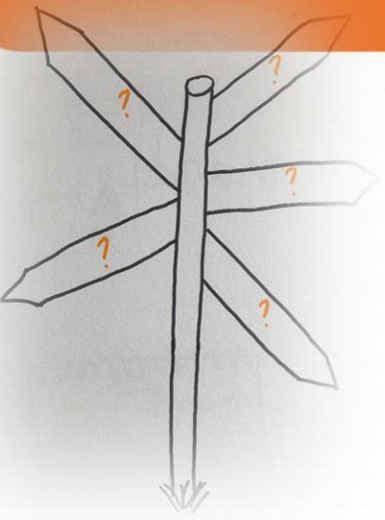




Articulation enseignement/diffusion et recherche



- L'enseignement m'a amenée aux études sur le genre
- Les études sur le genre m'ont amenée à :
 - Des questionnements importants autour des catégories
 - Une exploration pluridisciplinaire du point de vue thématique (merci à la centaine de mémoires en étude sur le genre !)
 - La notion de savoirs situés
 - De la réflexivité
- Vers de nouveaux objets de recherche avec la convergence de l'informatique et des études sur le genre
- Construire de la transmission autour de ces sujets
 - Semaine intensive sur « genre et numérique » pour les étudiant.es en études sur le genre
 - Divers cours transversaux sur genre et informatique
 - Diverses interventions en dehors de l'université



Faire des choix



- Richesse du métier... : nécessité de prioriser, de choisir, de reNONcer
→ réflexions ont abouti à une grille d'auto-questionnement

Question générale	Question plus spécifique	OUI	NON	A VOIR
Ai-je la capacité de m'investir ?	Ai-je le temps matériel ?			
Est-ce que cela fait sens pour moi ?	Est-ce en adéquation avec mes valeurs, mon éthique ?			
	Est-ce que cette opportunité est utile à la société ?			
	Au service du plus grand nombre ?			
	Au service de personnes en situation de vulnérabilité ?			
	Est-ce qu'il y a une entente avec les personnes concernées ? L'envie de collaborer est-elle présente ?			
	Est-ce que ce projet implique un travail croisant d'autres disciplines que l'informatique ?			
	Est-ce que j'ai l'élan ? Est-ce que je me sens à ma place ?			
	Est-ce que c'est un projet auquel je pense pouvoir contribuer ?			



Regarder devant :
programme de recherche



Perspectives : pour qui, quoi ?

- Pour une recherche réflexive : contributions à la scientométrie
 - ☐ Poursuivre les réflexions autour de l'articulation de la recherche et des enjeux écologiques
 - ☐ Poursuivre le travail en lien avec les politiques d'égalité
 - ☐ Permettre l'analyse des données de la science selon des perspectives diverses : lacs de données pour la scientométrie
 - ☐ Contribuer à garantir la qualité de la production scientifique avec la détection de faux articles ou de problèmes à partir des images

- Pour une recherche inclusive
 - ☐ Langues et biais : modèles de langue et traductions automatiques
 - ☐ Informatique décisionnelle inclusive accessible, notamment en lien avec le handicap visuel
 - ☐ Pédagogie et place de l'humain dans les plateformes éducatives



Enjeux de garantir la qualité de la recherche scientifique



Science 16/02/2024 17:19 | Actualisé le 17/02/2024 14:29

Intelligence artificielle : l'image absurde d'un rat au pénis XXL sème le trouble chez les scientifiques

La faute est grossière pour une étude scientifique, et pose un vrai problème.

Par Matthieu Balu

INTELLIGENCE ARTIFICIELLE - Ce serait drôle (et juste un tout petit peu gênant) si ce n'était pas la surface émergente d'un gros problème pour la communauté scientifique mondiale. Le 13 février dernier, des résultats de recherches sur la fertilité ont été publiés dans une revue spécialisée, *Frontiers in Cell Development and Biology*. Une étude tout à fait sérieuse, validée par une publication reconnue... Mais qui visiblement n'a pas regardé de près les illustrations accompagnant la recherche.

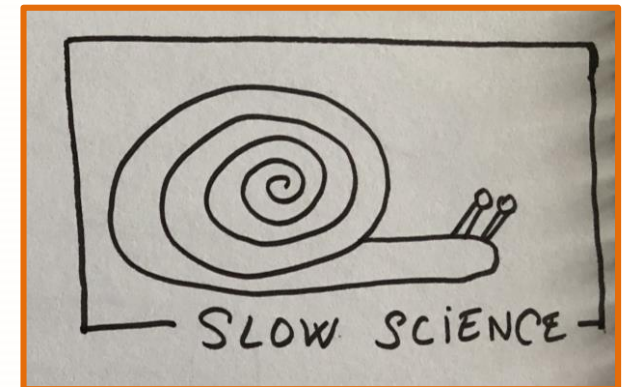
Il faut dire que cette anecdote pose de manière imagée un problème auquel le monde scientifique est confronté d'une manière aussi aiguë, sinon plus, qu'ailleurs. L'arrivée massive des images créées par IA doit être régulée, et les revues scientifiques n'ont pas toutes la même politique... Ni la même acuité à les repérer. Alors que la revue *Nature* a déjà fait savoir qu'elle n'accepterait aucun contenu créé par IA, les *Frontiers in* (qui sont tout un catalogue de revues spécialisées par domaine) le permettent, si la nature de l'image est clairement partagée avec les relecteurs et mentionné sur l'étude.

Le cas de notre rat virilisé montre surtout que cette règle n'a guère fonctionné, faute d'avoir porté suffisamment attention au matériel accompagnant l'étude. Une faute qui peut écorner sensiblement la crédibilité de ces journaux scientifiques : si ces derniers, dont l'exactitude et la rigueur sont le cœur de métier, laissent passer de tels manquements, que faut-il attendre des études qu'ils publient elles-mêmes ? Pas de réponse de *Frontiers in Cell Development and Biology*, mais la mention en haut de l'étude qu'une correction est en cours...



Enjeux de garantir la qualité de la recherche scientifique

- Dimension systémique du fonctionnement de la recherche imparfait
- Enjeux de qualité qui s'oppose à la rapidité (quantité)
- Revenir aux notions de fiabilité, d'éthique, d'intégrité scientifique, de responsabilité scientifique... vers une « science durable »
- Travaux sur l'analyse des images d'articles





Inclusion sur les enjeux de l'importance visuelle

■ Application **BeMyEyes**

- Lancée en 2015 comme une plateforme d'entraide entre personnes non voyantes ou malvoyantes et volontaires voyantes
- Accueil en 2023 d'un « nouveau bénévole » disponible à n'importe quelle heure du jour ou de la nuit : GPT-4V. → Description du visuel envoyé
- Fonctionnement :
 - Nouvel onglet **Be My AI** dans l'application avec prise de photo de l'objet ou de la scène à décrire
 - Photo envoyée sur les serveurs d'OpenAI pour être analysée par GPT-4V (Vision)
 - Dizaine de secondes plus tard : description affichée dans l'application et possibilité de lecture par VoiceOver (lecteur d'écran intégré)



Inclusion sur les enjeux de l'importance visuelle

■ Be My AI :

Il s'agit d'un dessin fait à la main avec des éléments écrits et illustrés. En haut, il y a un titre "HABILITATION À DIRIGER DES RECHERCHES" avec un calendrier de mars 2021 à côté, où le jeudi 4 mars 2021 est encerclé.

En dessous, il y a une figure humaine stylisée avec une robe à rayures orange et blanches, sans visage, tendant la main vers un papillon orange.

À droite de la figure, il y a un panneau avec le texte "En présentiel à Lyon et en visio" et en dessous "De l'analyse informatique de données de la société à l'analyse sociale de l'informatique".

En bas, le nom "Cécile Favre" est écrit, suivi de "Université Lumière Lyon 2" et "Laboratoire ERIC".





Inclusion sur les enjeux de l'importance visuelle

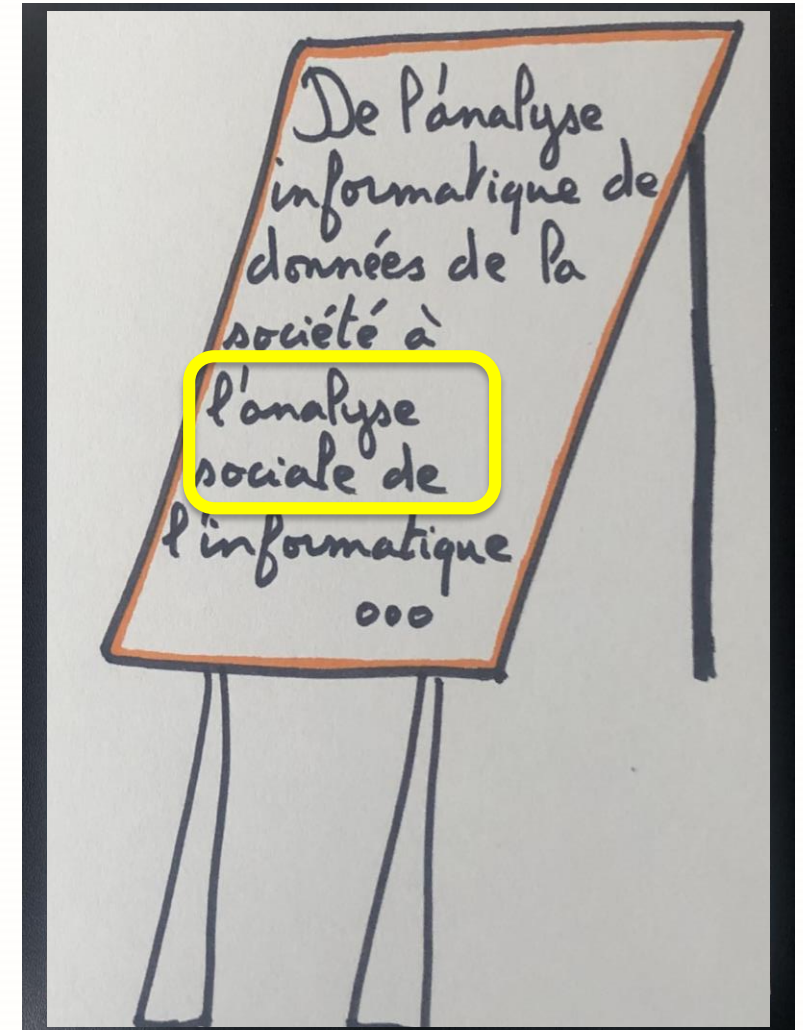
■ Be My AI :

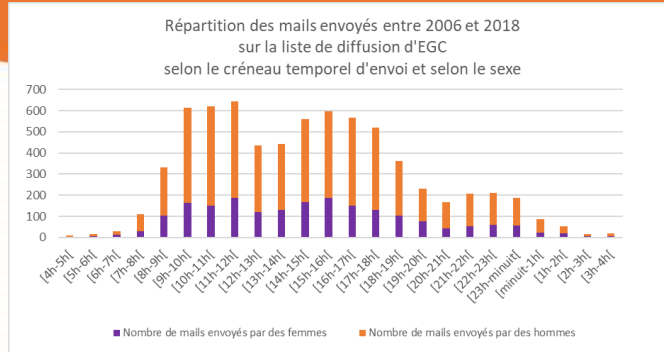
Il s'agit d'un dessin fait à la main avec des éléments écrits et illustrés. En haut, il y a un titre "HABILITATION À DIRIGER DES RECHERCHES" avec un calendrier de mars 2021 à côté, où le jeudi 4 mars 2021 est encerclé.

En dessous, il y a une figure humaine stylisée avec une robe à rayures orange et blanches, sans visage, tendant la main vers un papillon orange.

À droite de la figure, il y a un panneau avec le texte "En présentiel à Lyon et en visio" et en dessous "De l'analyse informatique de données de la société à l'informatique".

En bas, le nom "Cécile Favre" est écrit, suivi de "Université Lumière Lyon 2" et "Laboratoire ERIC".





Enjeux de l'accessibilité en informatique décisionnelle

■ Description de Be My AI :

Il s'agit d'un graphique à barres illustrant la répartition des mails envoyés entre 2006 et 2018 sur la liste de diffusion d'EGC selon le créneau temporel d'envoi et selon le sexe. Les barres violettes représentent le nombre de mails envoyés par des femmes et les barres oranges représentent le nombre de mails envoyés par des hommes. Les créneaux horaires sont indiqués en bas du graphique, allant de "0h-1h" à "23h-24h". Les valeurs sur l'axe vertical commencent à 0 et montent jusqu'à 700 avec des intervalles de 100.

Les barres les plus hautes se situent dans les créneaux "9h-10h" et "10h-11h" pour les hommes, avec des valeurs proches de 600 mails envoyés. Pour les femmes, les barres les plus hautes se trouvent également dans les créneaux "9h-10h" et "10h-11h", mais les valeurs sont légèrement inférieures à celles des hommes, autour de 500 mails envoyés.

Globalement, les hommes envoient plus de mails que les femmes dans presque tous les créneaux horaires, à l'exception de "3h-4h" et "4h-5h" où les femmes envoient légèrement plus de mails.

Enjeux sur les biais de traduction

Extrait d'atelier
sur Google Translate :

Hongrois :
neutralité
des pronoms
personnels



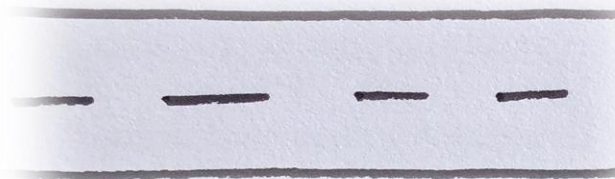
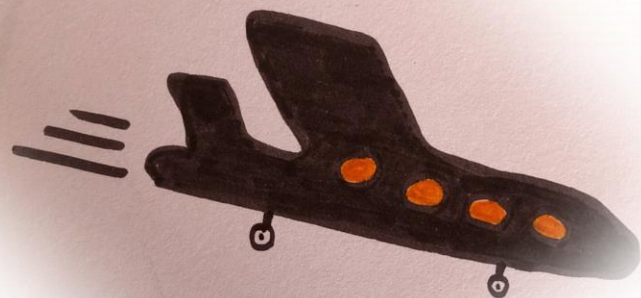
Double traduction
en anglais



Traduction en
français porteuse de
stéréotypes



Français	Français -> Hongrois	Hongrois -> Anglais	Hongrois -> Français
Elle fait la cuisine	Ő főz	She/he cooks	Elle cuisine
Il fait la cuisine	Ő főz	She/he cooks	Elle cuisine
Il aime bricoler	Szeret bütykölni	She/He likes to tinker	Il aime bricoler
Elle aime bricoler	Szeret bütykölni	She/He likes to tinker	Il aime bricoler
J'ai rencontré un informaticien passionné	Találkoztam egy szenvedélyes informatikussal	I met a passionate computer scientist	J'ai rencontré un informaticien passionné
J'ai rencontré une informaticienne passionnée	Találkoztam egy szenvedélyes informatikussal	I met a passionate computer scientist	J'ai rencontré un informaticien passionné



Pour conclure



Pour conclure

■ Une HDR façon « manifeste »

- « **Écrit public** par lequel un chef d'État, un gouvernement, un parti, etc., rend compte de son mandat ou expose son programme, son point de vue sur un problème politique »
- « Exposé théorique par lequel des écrivains, des artistes lancent **un nouveau mouvement** »



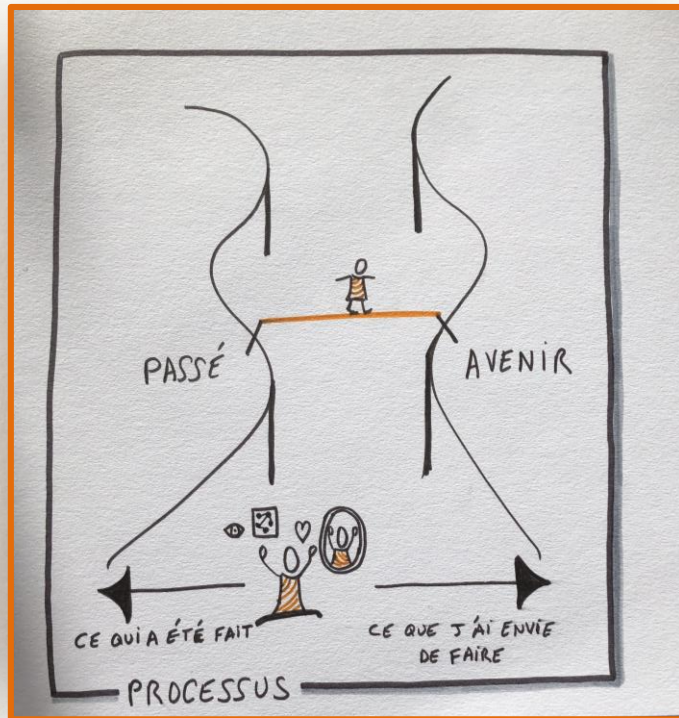
■ Et aussi...

- Un processus de bilan professionnel, et personnel à un certain niveau
- Un processus de conscientisation
 - De mon identité scientifique
 - De ma posture de chercheuse, d'accompagnante, d'enseignante, ...
- Un processus pour penser la suite à long terme
 - Quoi ?
 - Comment et avec qui ?
- Un aboutissement : une occasion d'expression d'un contenu qui voudrait **inviter à la réflexion, à la réflexivité, à la discussion**

Pour conclure

■ Une HDR façon « manifeste »

- « **Écrit public** par lequel un chef d'État, un gouvernement, un parti, etc., rend compte de son mandat ou expose son programme, son point de vue sur un problème politique »
- « Exposé théorique par lequel des écrivains, des artistes lancent **un nouveau mouvement** »



■ Et aussi...

- Un processus de bilan professionnel, et personnel à un certain niveau
- Un processus de conscientisation
 - De mon identité scientifique
 - De ma posture de chercheuse, d'accompagnante, d'enseignante, ...
- Un processus pour penser la suite à long terme
 - Quoi ?
 - Comment et avec qui ?
- Un aboutissement : une occasion d'expression d'un contenu qui voudrait **inviter à la réflexion, à la réflexivité, à la discussion**



Pour conclure : une invitation à la réflexion



Dominique Pestre, historien des sciences,
dans *Introduction aux Sciences studies* (2006), p 94.

« Malgré ce qu'énonce la philosophie spontanée des savants, il est difficile de laisser la science dans sa tour d'ivoire et de faire comme si c'était dans l'univers clos des laboratoires que se passait l'essentiel de ce qui la concerne. [...], la question du rapport des sciences [...] à l'économique et au politique (comme aux sociabilités et à la domination masculine) ne peut être évacuée. »

Lyon, 7 mars 2024



UFR **anthropologie**
SOCIOLOGIE
SCIENCE POLITIQUE



De l'analyse informatique
de données de la société
à l'analyse sociale de l'informatique.

Un cheminement guidé par les études de genre
vers un décloisonnement disciplinaire
et une posture réflexive.

Soutenance pour l'obtention de
l'Habilitation à Diriger des Recherches

