

Votre méthode de conduite de la recherche

N.Mandran (LIG Grenoble)

Octobre 2020

Mandran, N. (2018). *Traceable human experiment design research: Theoretical model and practical guide* (ISTE, Vol. 9). Wiley.

Mandran, N., & Dupuy-Chessa, S. (2017). THEDRE: A Traceable Process for High Quality in Human Centred Computer Science Research. *International Conference on Information Systems Development (ISD)*.

<https://aisel.aisnet.org/isd2014/proceedings2017/ISDMethodologies/12>

Mandran, N., & Dupuy-Chessa, S. (2018). *Supporting experimental methods in information system research*. 1–12. <https://doi.org/10.1109/RCIS.2018.8406654>



Pour citer ce document : N.Mandran, Octobre 2020, « Votre méthode de conduite de la recherche ». Formation du collège des écoles doctorales de Grenoble.



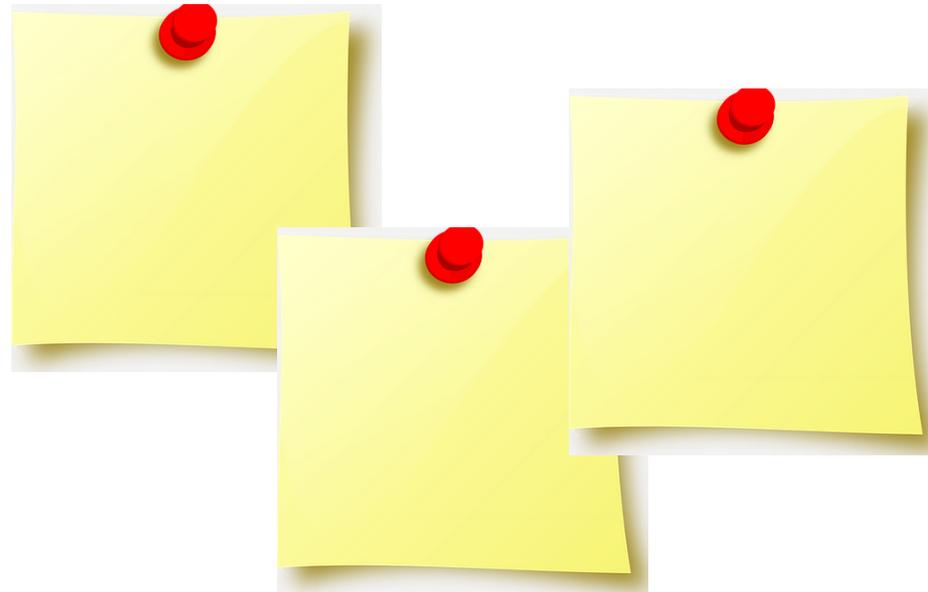
Quelques mots de présentation ...

- Nadine Mandran, Méthodologue
 - Statistiques, SHS, Démarche qualité
 - Ingénieur encadrant des doctorants sur les méthodes de production et analyse de données
 - Spécialisée dans les méthodes de conduite de la recherche en informatique centrée humain – THEDRE



Brainstorming

Vos questions ou difficultés par rapport à ce travail de thèse ?



Une thèse



Une thèse

- à identifier un sujet,
- à rassembler des documents (données) sur ce sujet
- à les analyser, les organiser,
- à construire une proposition, à l'évaluer
- se faire comprendre, défendre ses idées



Aussi

- on oubliera ainsi de parler comme un poète
- pas non plus obligé de tout lire sur le sujet mais le cœur de cible

Une thèse est une expérience de travail, mais c'est aussi un pari, un jeu, une chasse au trésor.

Le vrai défi dans une thèse, c'est de la vivre comme un défi, de se mettre soi-même au défi.





Seulement 3 ans pour tout ça

Une thèse

- à identifier un sujet, // identify a problematic
- à rassembler des documents (données) sur ce sujet //gathered data and documents
- à les analyser, les organiser, // analyze and manage them
- à construire une proposition, à l'évaluer // build a contribution and test it
- se faire comprendre, défendre ses idées //make yourself understood,defend your ideas

LIRE, CONCEVOIR, IMAGINER, DEVELOPPER, TESTER, REDIGER, REDIGER, REDIGER, RELIRE, RECONCEVOIR, AMELIORER, RETESTER, REDIGER, REDIGER, REDIGER, ...

READ, DESIGN, IMAGINE, DEVELOP, TEST, WRITE,WRITE,WRITE, REREAD, REDESIGN, ENHANCE, RETEST, WRITE,WRITE,WRITE





Seulement 3 ans // Only 3 years

Une thèse // Thesis

- à identifier un sujet, // identify a problematic
- à rassembler des documents (données) sur ce sujet // gathered data and documents
- à les organiser, // analyze and manage them
- à construire une proposition, à l'évaluer // build a contribution and test it
- se faire comprendre, défendre ses idées // make yourself understood, defend your ideas

READ, DESIGN, IMAGINE, DEVELOP, TEST, WRITE, WRITE, WRITE, REREAD, REDESIGN, ENHANCE, RETEST, WRITE, WRITE, WRITE

The real challenge in a thesis is to live it as a challenge, to challenge oneself.



PILOTER VOTRE THESE // DRIVE YOUR THESIS

AVOIR DES GUIDES ET DES OUTILS // GUIDES AND TOOLS





Objectifs de ces deux jours

- Concevoir **votre** méthode de conduite de la recherche dans un cadre de qualité
- Avoir des outils pour piloter votre thèse
 - Bien commencer ?
 - Problématique ?
 - Séquencer le travail de thèse ?
 - Etat de l'art ?
 - Protocole expérimental ? Production de données ?
 - Analyser des données quantitatives ?



Avant propos ...

Méthodologie ou méthode ?

- Méthodologie : étude de la méthode, conception de méthodes

« **Methodology** is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge. Typically, it encompasses concepts such as paradigm, theoretical model, phases and quantitative or qualitative techniques » (Berg, 2009).

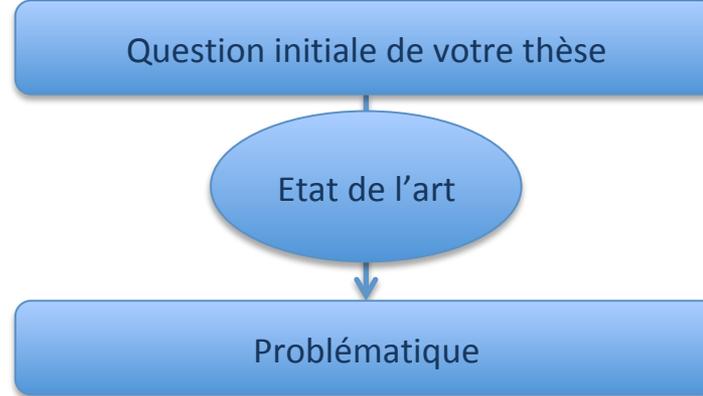
- Méthode : ensemble de tâches, d'outils, de procédures assemblés et organisés dans le temps pour atteindre un but.

Method is “an integrated collection of procedures, techniques, product descriptions, and tools, for effective, efficient, and consistent support of the engineering process” (Harmsen et al., 1997).

- **Nous utilisons uniquement le terme méthode que nous qualifions par d'autres termes : méthode de conduite de la recherche, méthode de production de données, méthode d'analyse de données ...**

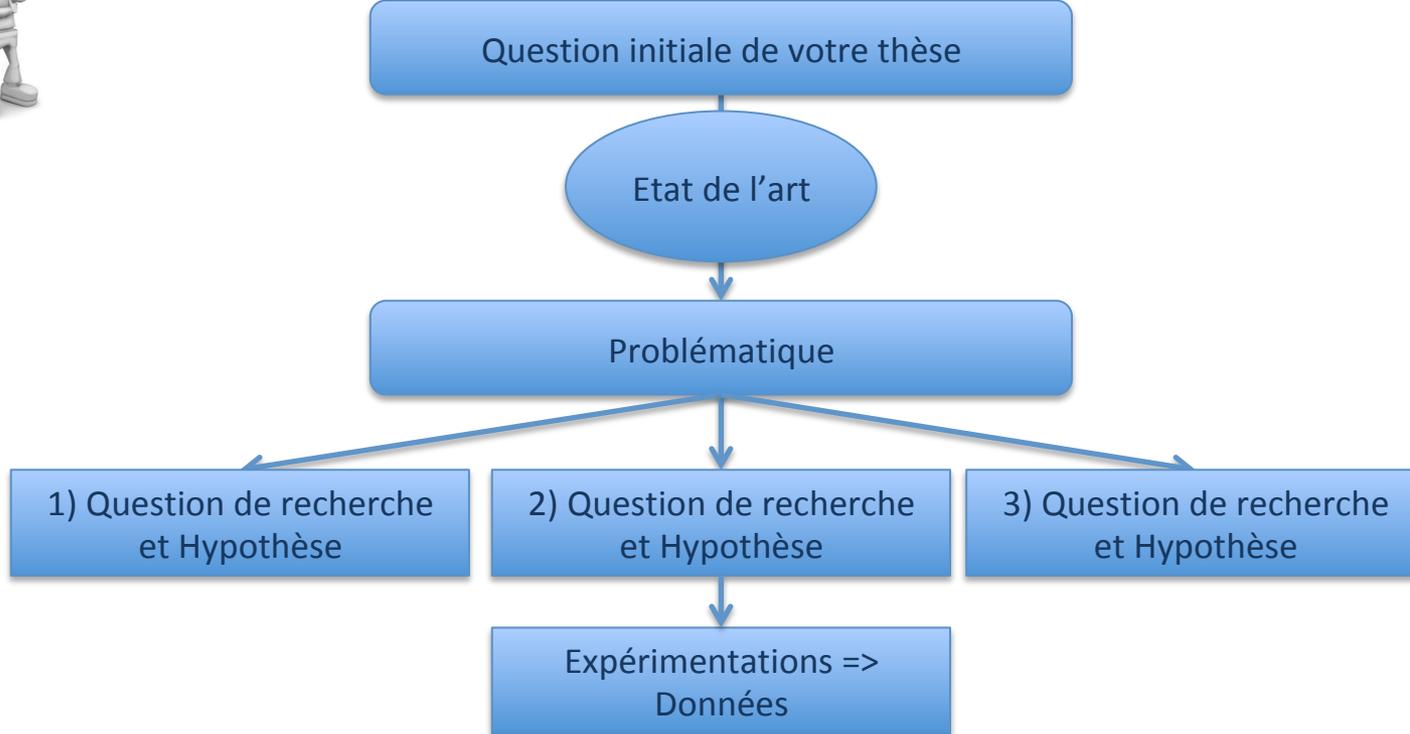


Une vision d'ensemble



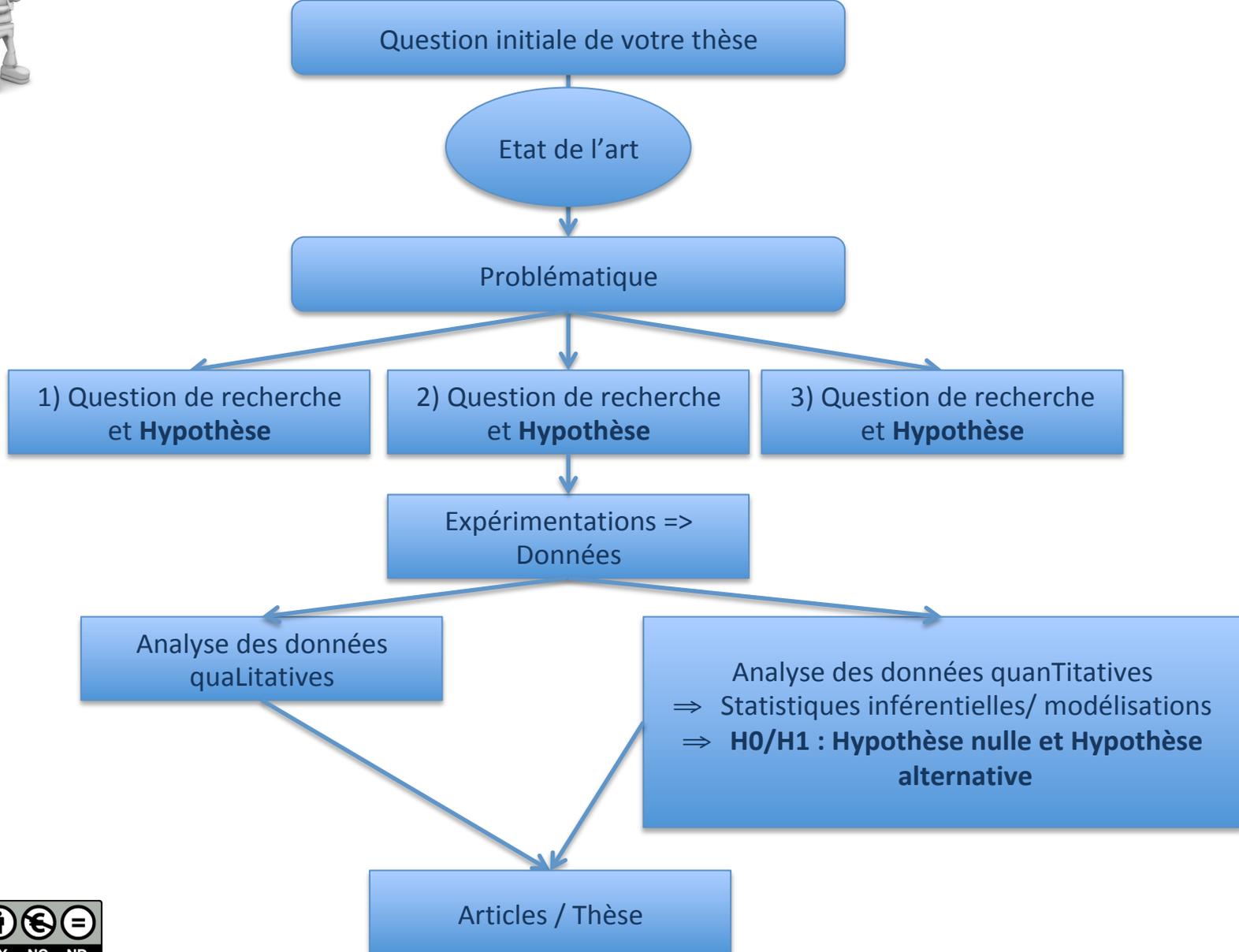


Une vision d'ensemble



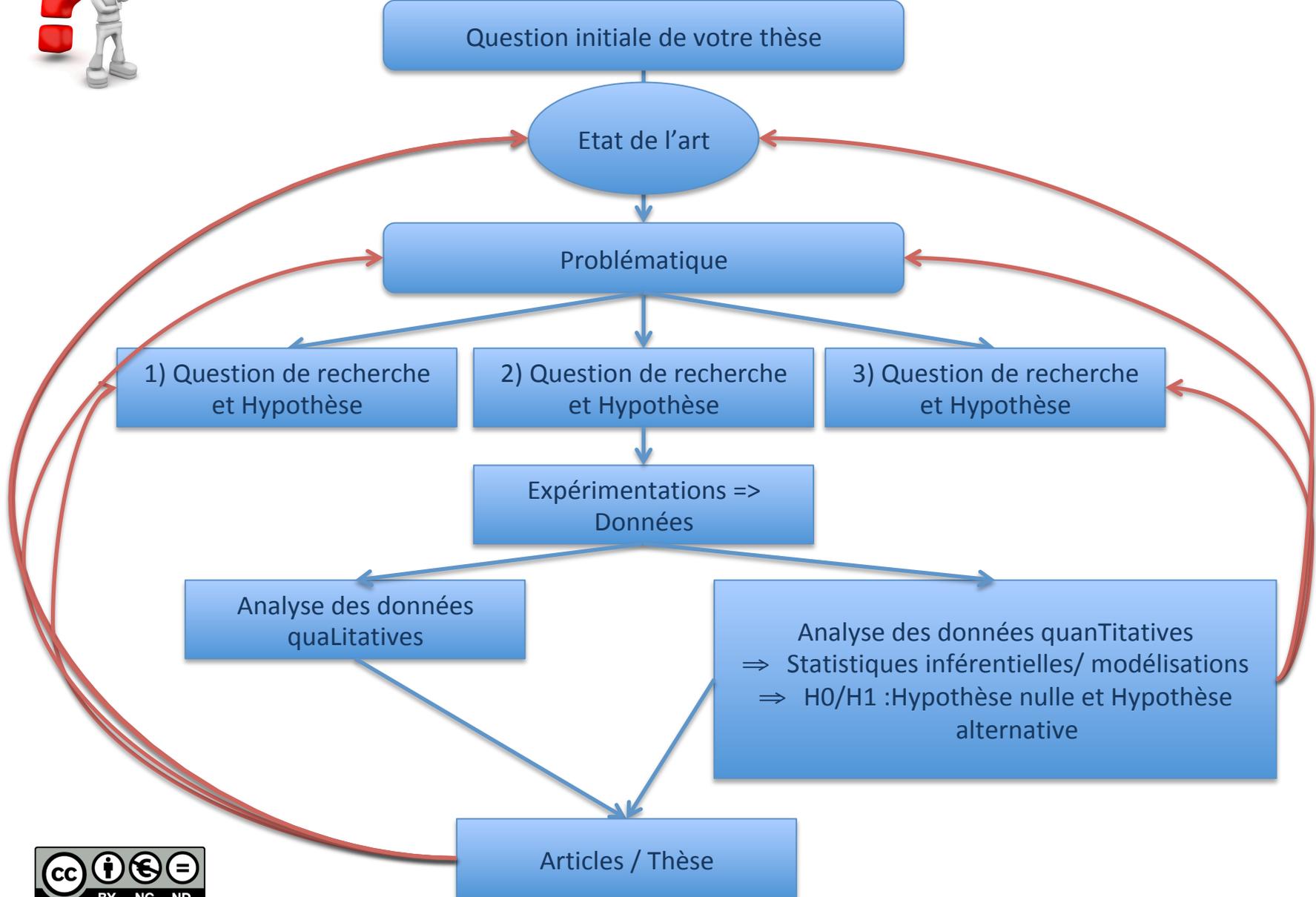


Une vision d'ensemble





Une vision d'ensemble





La thèse un projet ..

- Que va produire se projet ?
- Quelles seront vos contributions à la fin de votre travail de doctorat ?





Exemple : simulateur TELEOS

ANR TELEOS, V.LUENGO ET AL.

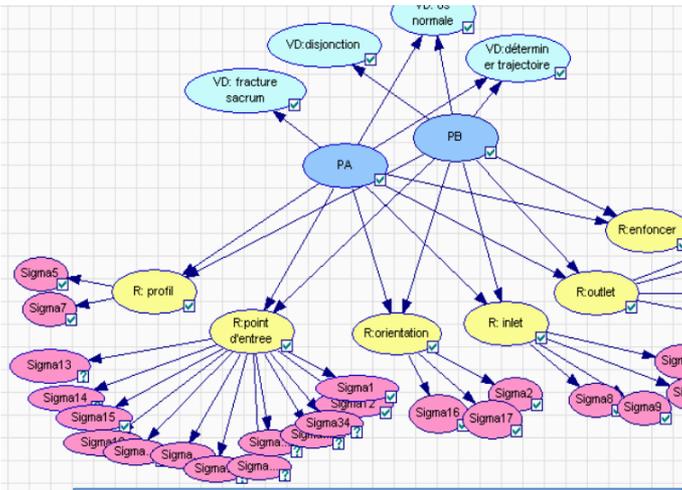


Simulateur pour l'apprentissage de la chirurgie percutanée :

L'outil activable produit par la recherche



Exemple : simulateur Teleos



Modèle de l'apprenant



Le simulateur



Exemple : simulateur TELEOS

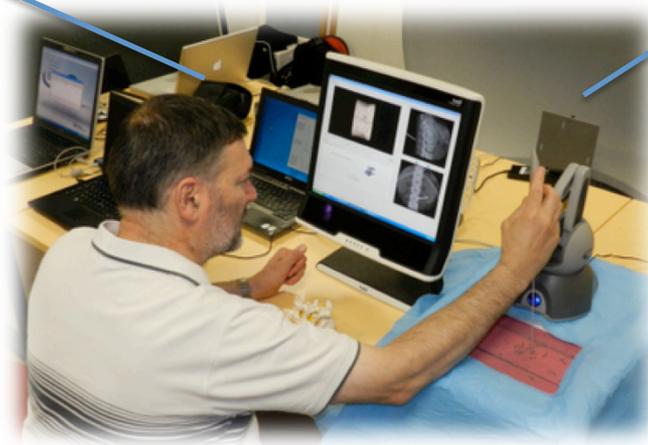
L'outil activable est décomposable en sous-parties

L'interface :

- Terminologie
- Fonctionnalités
- ...

Le bras haptique :

- Manipulation
- Résistance
- ...





Exemple : simulateur Teleos

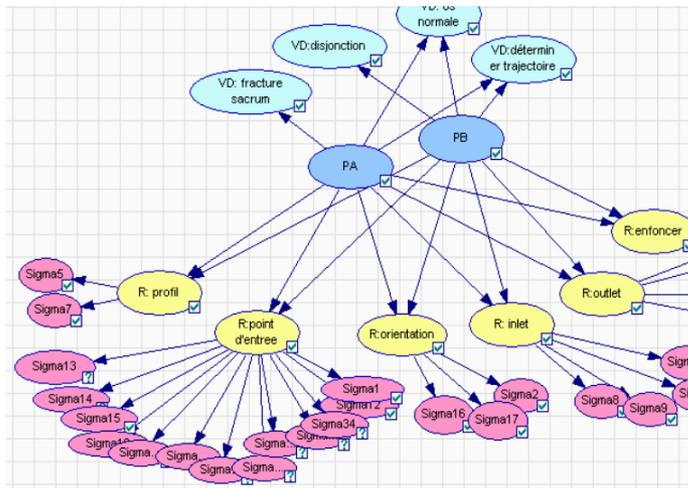
Composants Activables (CA)

L'interface :

- Terminologie
- Fonctionnalités
- ...

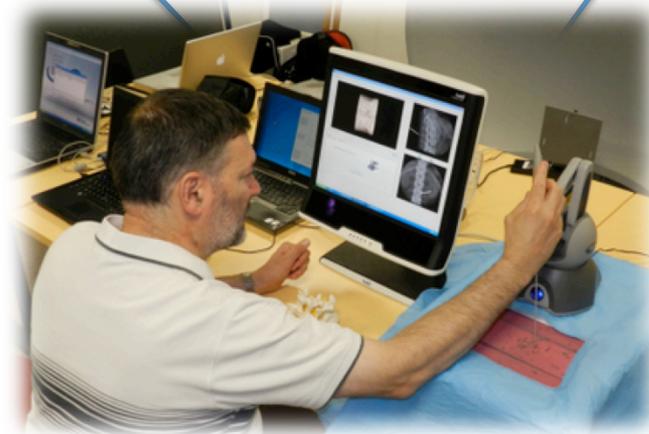
Le bras haptique :

- Manipulation
- Résistance
- ...



Modèle de l'apprenant

Connaissances scientifiques (KS)



Le simulateur

Outils Activables (OA)





Exemple : simulateur Teleos



T1 : In situ

OA n'existe pas



T2 : Maquettage

OA existe statique



T3 : Simulateur

OA existe dynamique

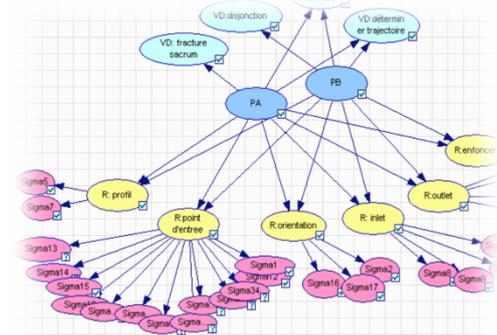


Figure 2 - le réseau des conceptions



Exemple : simulateur Teleos



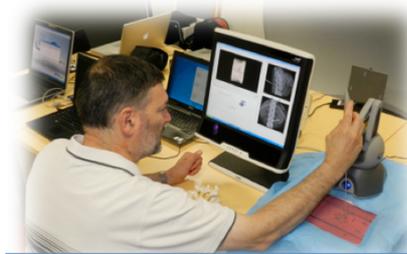
T1 : In situ

OA n'existe pas



T2 : Maquettage

OA existe statique



T3 : Simulateur

OA existe dynamique

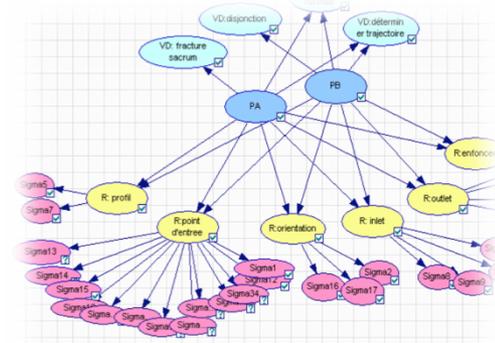


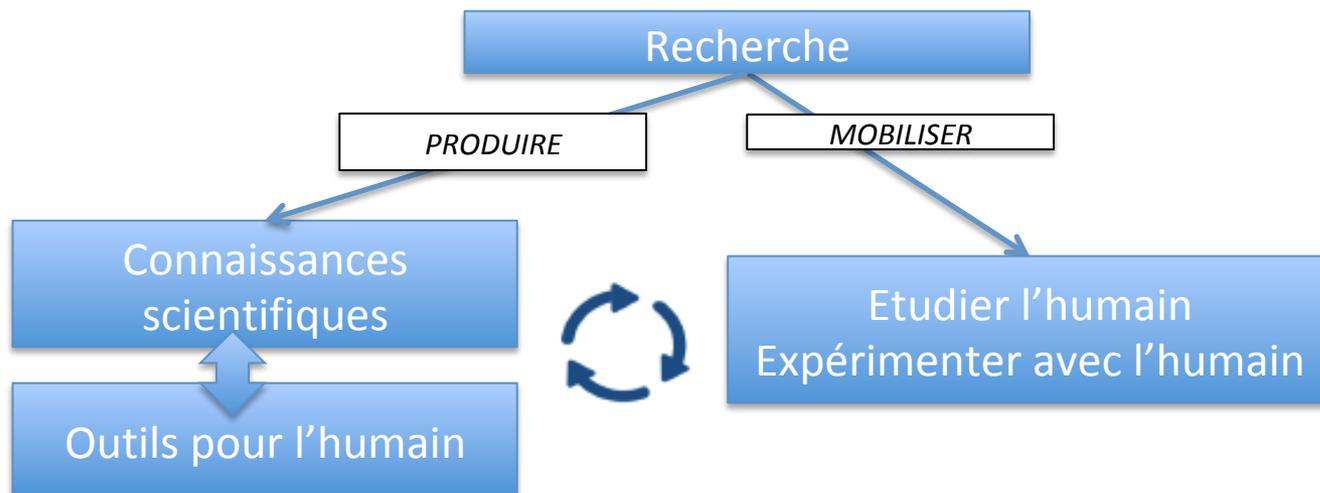
Figure 2 - le réseau des conceptions





L'ancrage de notre méthode de conduite de la Recherche en Informatique Centrée Humain

- **Un cadrage scientifique global, les sciences de l'artificiel (H.Simon 2004)**
 - **Un outil** que l'humain va pouvoir mobiliser : un EIAH, un dispositif d'enseignement, un Dashboard, ...
 - Des **connaissances scientifiques** basées sur l'activité humaine : modèle de l'apprenant, modèle de l'enseignant, modèle pédagogique, modèle de traces, ...



- Herbert A. Simon. 2004. *Les Sciences de l'artificiel* (traduction J.L. Lemoigne Éd. rev. et complétée ed.). Folio, Paris •



Contributions et outils

« Préciser les objectifs du travail de thèse et initier la collaboration entre doctorant et encadrant ... »

Quand : Ce document doit être utilisé pour introduire le travail de thèse au doctorant. C'est un prérequis pour utiliser le schéma cyclique de la méthode THEDRE.

Pourquoi : Pour présenter le type de contribution scientifique que le doctorant va produire et s'il doit développer des outils associés à cette contribution (ex. plateforme numérique, robot, etc.).

Qui : Les items 1 à 5 sont à préparer par le directeur de thèse. Ils sont ensuite à discuter entre doctorants et encadrants. Le doctorant peut les utiliser pour demander des précisions à son encadrant.

Apports : Pour clarifier ce qui est attendu du travail de thèse et présenter le domaine aux doctorants

Ce qu'en dise les doctorants : « *Je pense que ce guide peut faire gagner beaucoup de temps pour bien cibler ses objectifs.* » ; « *Approche intéressante, qui permet de subdiviser la perception d'une thèse en éléments compréhensibles.* »



Debriefing

- Types de contributions scientifiques ?
- Types d'outils , sont ils décomposables ?
- Et l'humain ?
- Incrémentale et Itérative ?



Discussion ...



Répétabilité ?
Reproductibilité ?

Protocole de production de données ?

Les données ?

Les résultats techniques ?

Le résultats scientifiques ?



Répétabilité vs Reproductibilité

- ISO 3534-1

- **Conditions de répétabilité**

- Conditions où les résultats d'essais indépendants sont obtenus par la même méthode sur des individus d'essai identiques dans le même laboratoire, par le même opérateur, utilisant le même équipement et pendant un court intervalle de temps.

- **Conditions de reproductibilité**

- Conditions où les résultats d'essai sont obtenus par la même méthode sur des individus d'essais identiques dans différents laboratoires, avec différents opérateurs et utilisant des équipements différents



Répétabilité vs Reproductibilité

- Quasiment impossible dans des autres sciences...
- Les protocoles de production des données sont reproductibles (?) :
 - Ils peuvent être réutilisés avec d'autres utilisateurs dans des conditions différentes ou similaires
 - Nécessite de tracer le processus de production et de traitement des données
 - Nécessité de tracer les évolutions de la production scientifique (connaissance scientifique et des outils)



- La répétabilité
- La reproductibilité
- La traçabilité

Dépend de

1. la posture épistémologique dans laquelle vous allez construire la connaissance scientifique
2. Les critères de valeur de validité que vous allez retenir pour votre travail



Cadre épistémologique

- Méthode de conduite de la recherche demande un positionnement épistémologique [Avenier & Thomas 2015]
- «L'épistémologie : étude de la constitution des connaissances valables» [Le Moigne 1995]
- Pour [Le Moigne 1995], cette définition pose des questions :
 - Qu'est-ce que la connaissance ?
 - Comment est-elle constituée ?
 - Comment apprécier sa valeur ou sa validité ?



Exemple d'une posture épistémologique

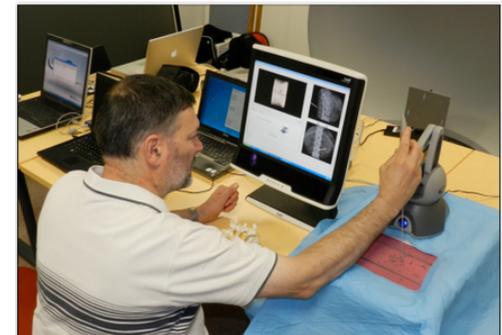
- **Constructivisme pragmatique** : [Avenier & Thomas 2015]
 - Hypothèse 1 : Des **représentations du réel** existent
 - Hypothèse 2 : **L'humain va exprimer ses connaissances du monde**
 - Hypothèse 3 : L'outil activable a **une finalité dans un contexte** donné
 - Hypothèse 4 : La construction de l'instrument est **incrémentale**
 - Hypothèse 5 : La **question de recherche peut évoluer**, en accord avec les contextes académique, technique et sociétal et les résultats du terrain





Exemple d'une posture épistémologique

- Valeur
 - Contribution : Incrémenter des connaissances existantes
 - Qualité du construit : Adéquation de l'outil activable aux activités humaines
- Validité
 - Mise à l'épreuve : Phase expérimentale où l'outil est construit et évalué
 - Multiplicité des données : Une large diversité des méthodes de production (qualitatif et quantitatif)
 - Fiabilité des données : Volume et qualité des données suivis par des indicateurs





Exemple d'une posture épistémologique

- **Constructivisme pragmatique** [Avenier & Thomas 2015]
 - Contribution : Incrémenter des connaissances existantes
 - Qualité du construit : Adéquation de l'outil activable aux activités humaines
 - Mise à l'épreuve : Phase expérimentale où l'outil est construit et évalué
 - Multiplicité des données : Une large diversité des méthodes de production (qualitatif et quantitatif) et de raisonnement
 - Fiabilité des données : Volume et qualité des données suivis par des indicateurs
 - **Documenter et tracer le travail de recherche** : protocole de production et de traitement des données, l'évolution de la connaissance et es outils activables



- **Votre posture épistémologique ?**
 - **Qu'est-ce que la connaissance dans votre domaine ?**
 - **Comment est-elle constituée ?**
 - **Comment apprécier sa valeur ou sa validité ?**



Traceable Human Experiment Design Research

THEDRE (Mandran 2018)

Pour

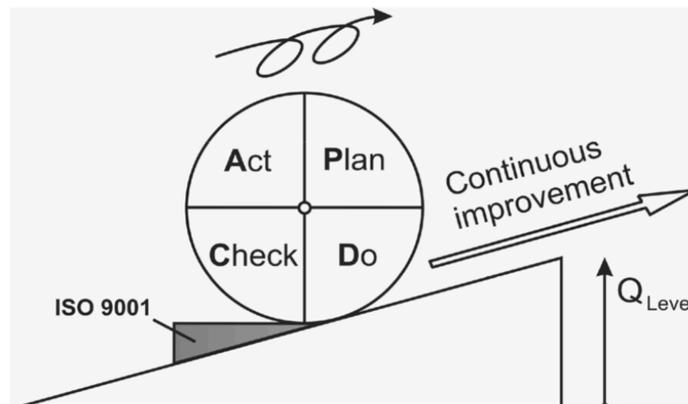
les recherches qui proposent des modèles de l'activité humaine en intégrant l'humain

des recherches qui développent des outils activables supports à l'activité humaine

+ de THEDRE Processus de guidage et outils

Avec un modèle de processus de conduite de la recherche ancré dans la démarche qualité et des indicateurs de traçabilité (Plan, Do, Check, Act – cycle de Deming)

Des guides pour accompagner l'utilisateur de THEDRE



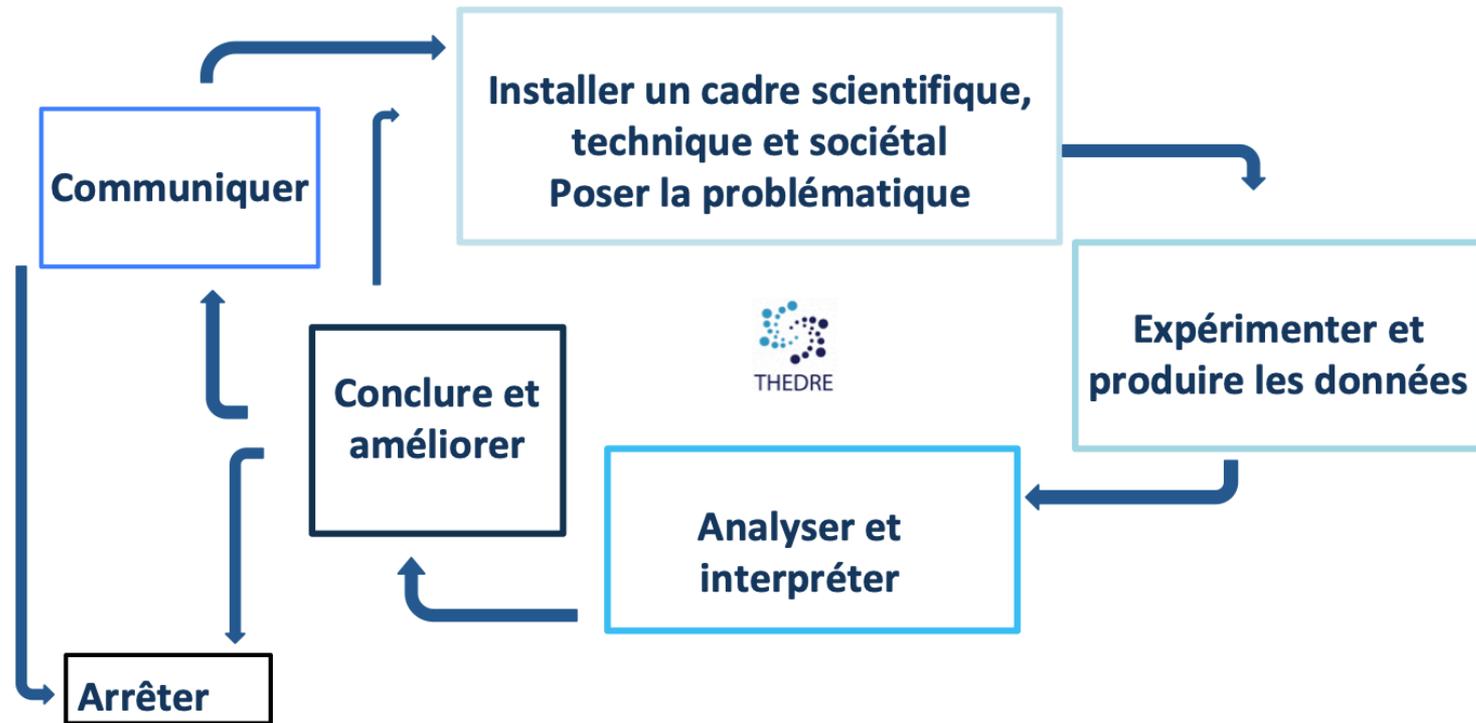
4 actions et un processus d'amélioration continue [Sokovic et al. 2010]





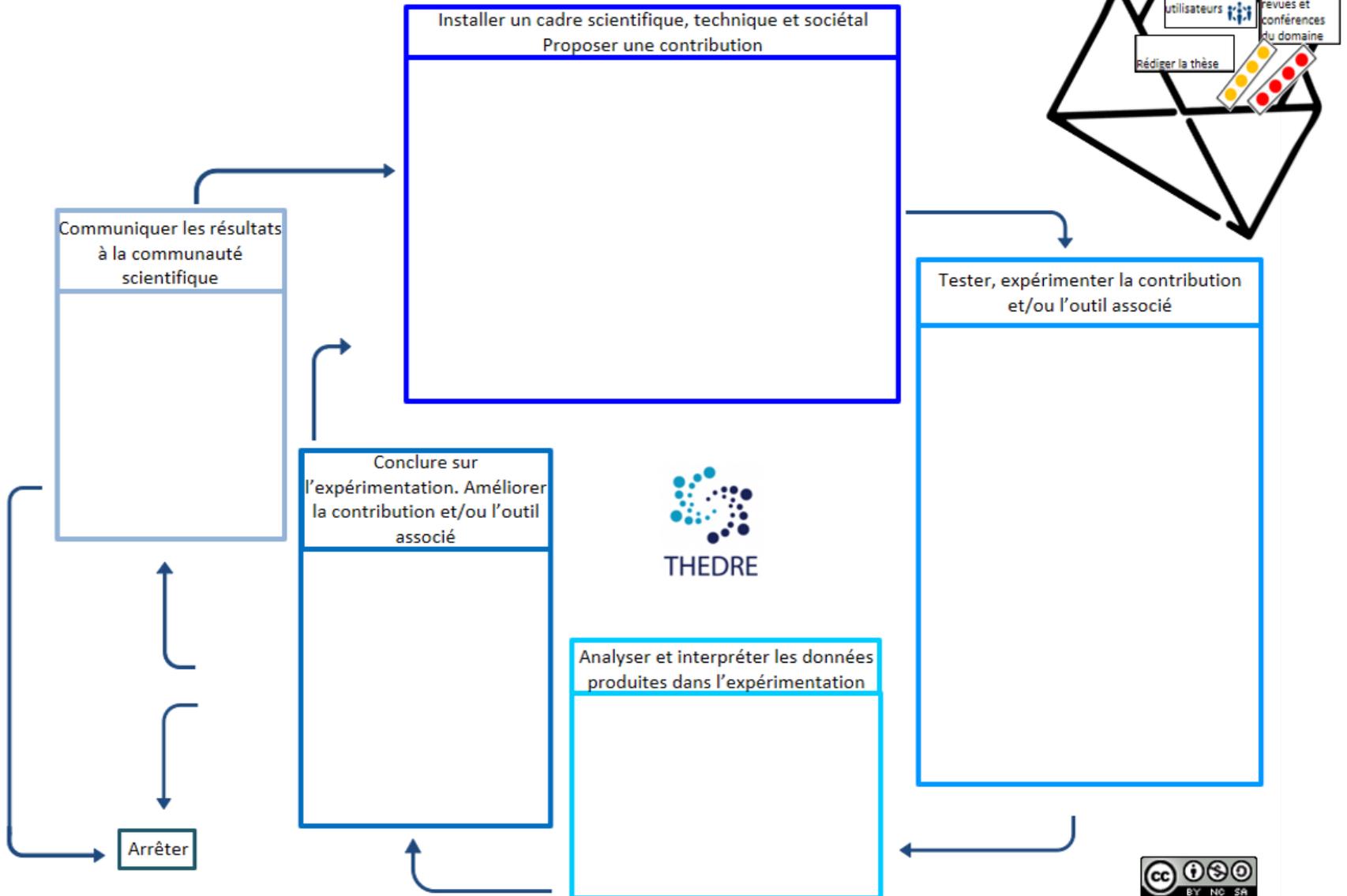


Une méthode adaptable





Votre méthode de conduite de la recherche itérative





Importance des tâches

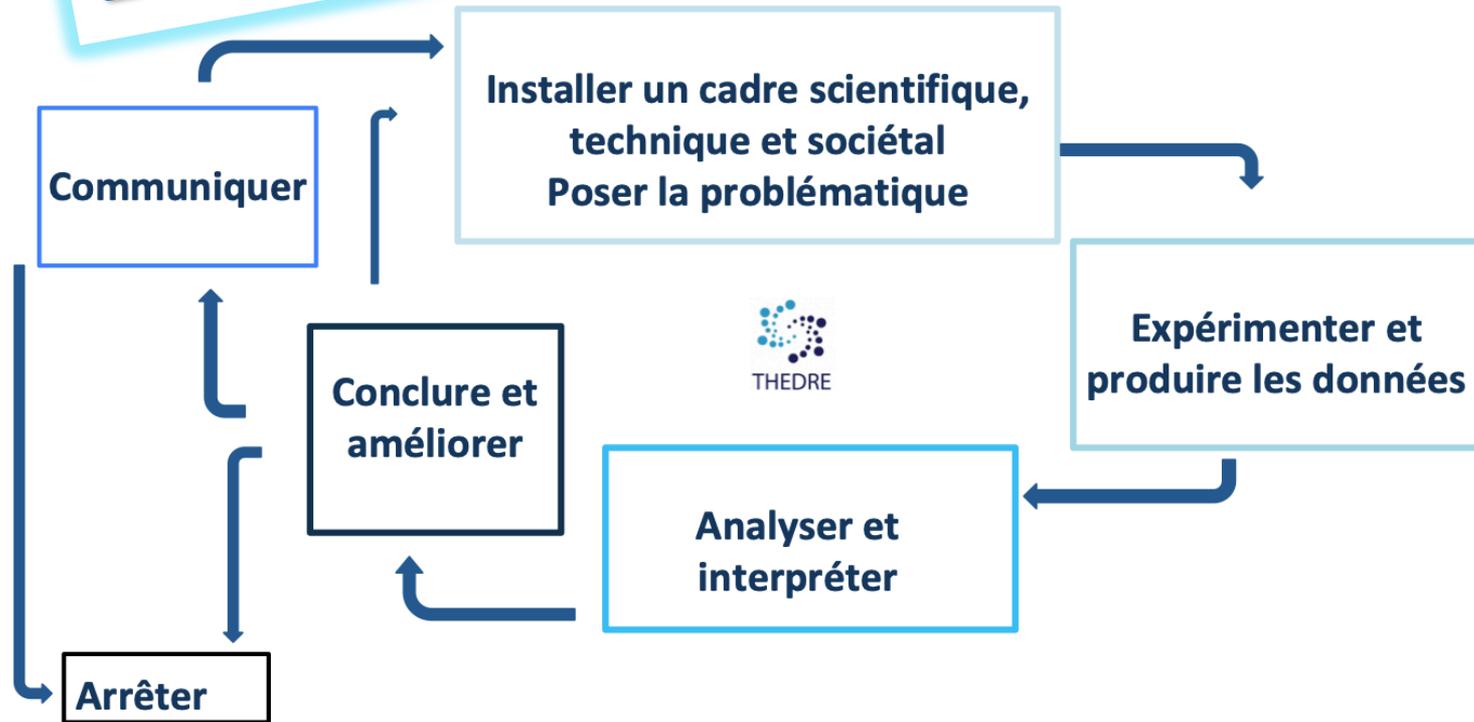
- Avec des gommettes rouge et orange venir indiquer les tâches qui vous semblent «bloquantes » et « à risque »





Une méthode adaptable

Debriefing







Guides et outils



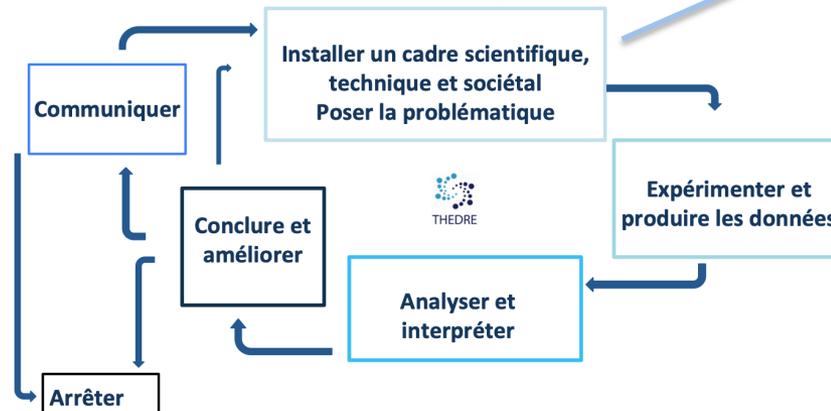
Pour bien commencer un travail de thèse et construire sa problématique



Spécifier la contribution scientifique et les outils associés



Spécifier les indicateurs et les mesures à prendre





Guides et outils



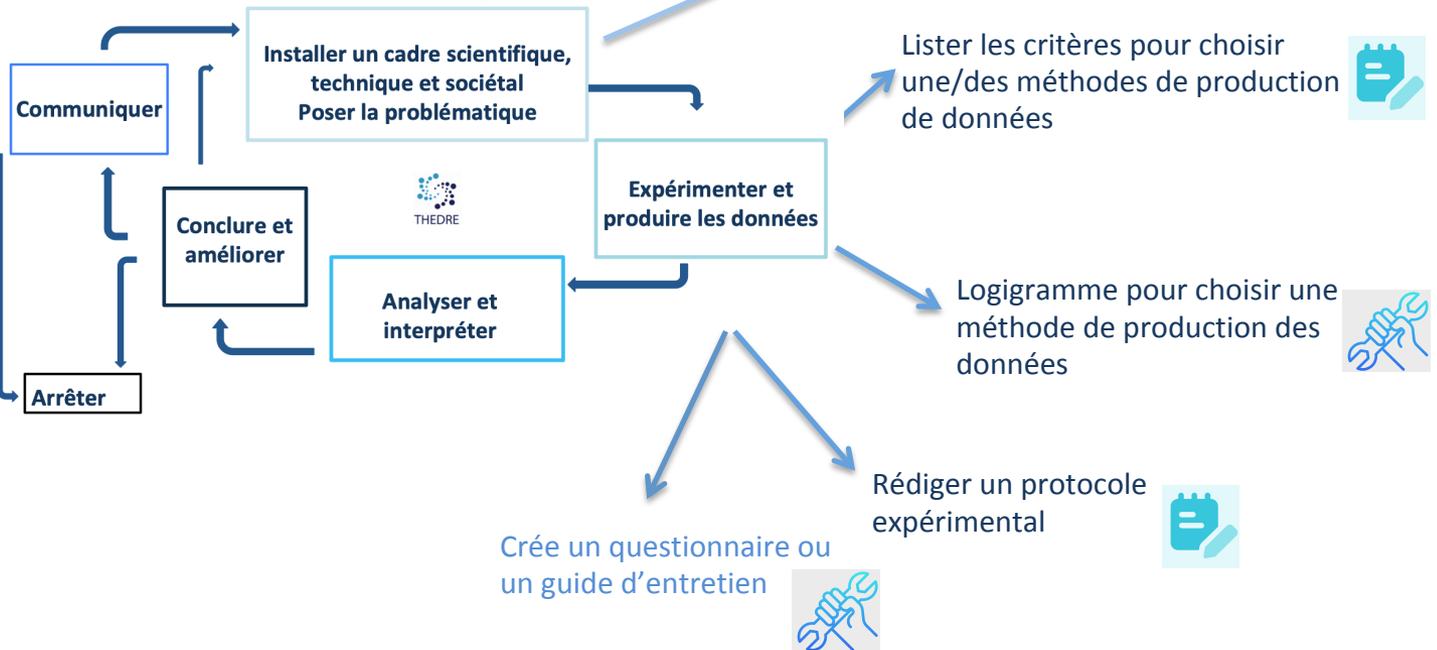
Pour bien commencer un travail de thèse et construire sa problématique



Spécifier la contribution scientifique et les outils associés



Spécifier les indicateurs et les mesures à prendre





Guides et outils



Pour bien commencer un travail de thèse et construire sa problématique



Spécifier la contribution scientifique et les outils associés



Spécifier les indicateurs et les mesures à prendre



Lister les critères pour choisir une/des méthodes de production de données



Logigramme pour choisir une méthode de production des données



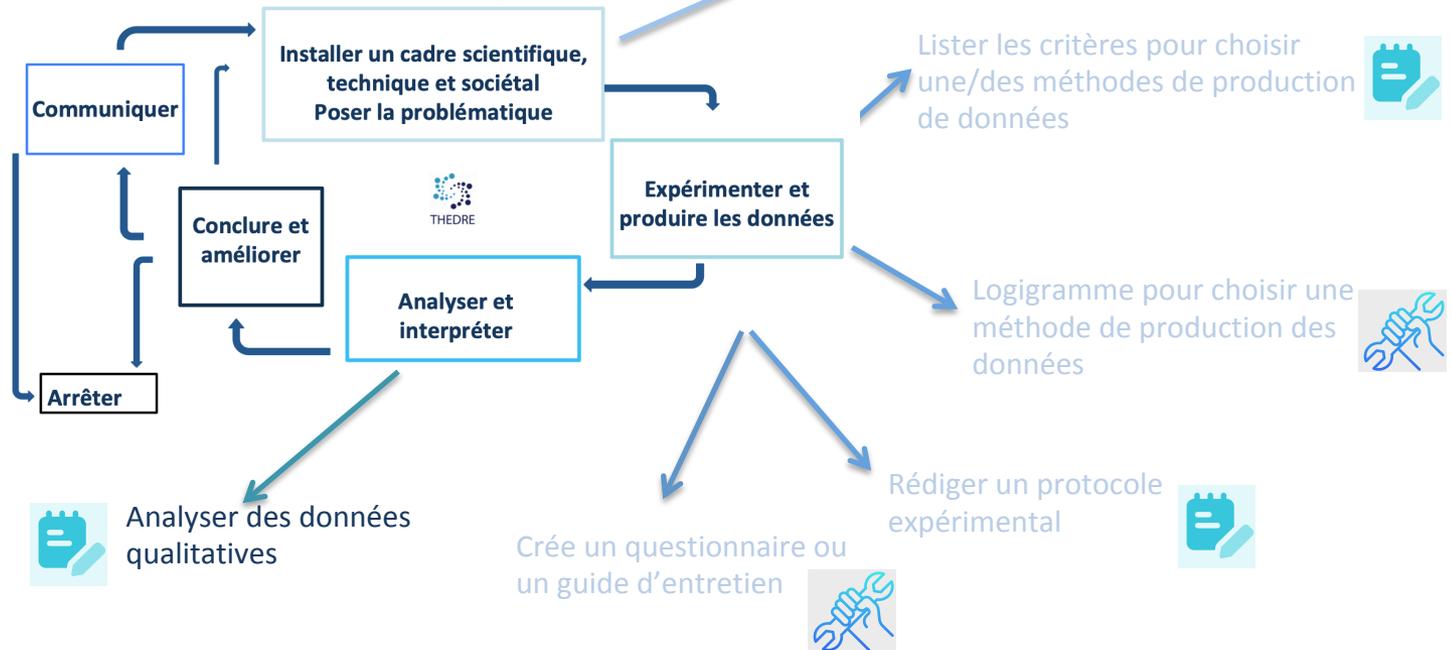
Rédiger un protocole expérimental



Crée un questionnaire ou un guide d'entretien



Analyser des données qualitatives





Guides et outils



Pour bien commencer un travail de thèse et construire sa problématique



Spécifier la contribution scientifique et les outils associés



Spécifier les indicateurs et les mesures à prendre



Lister les critères pour choisir une/des méthodes de production de données



Logigramme pour choisir une méthode de production des données



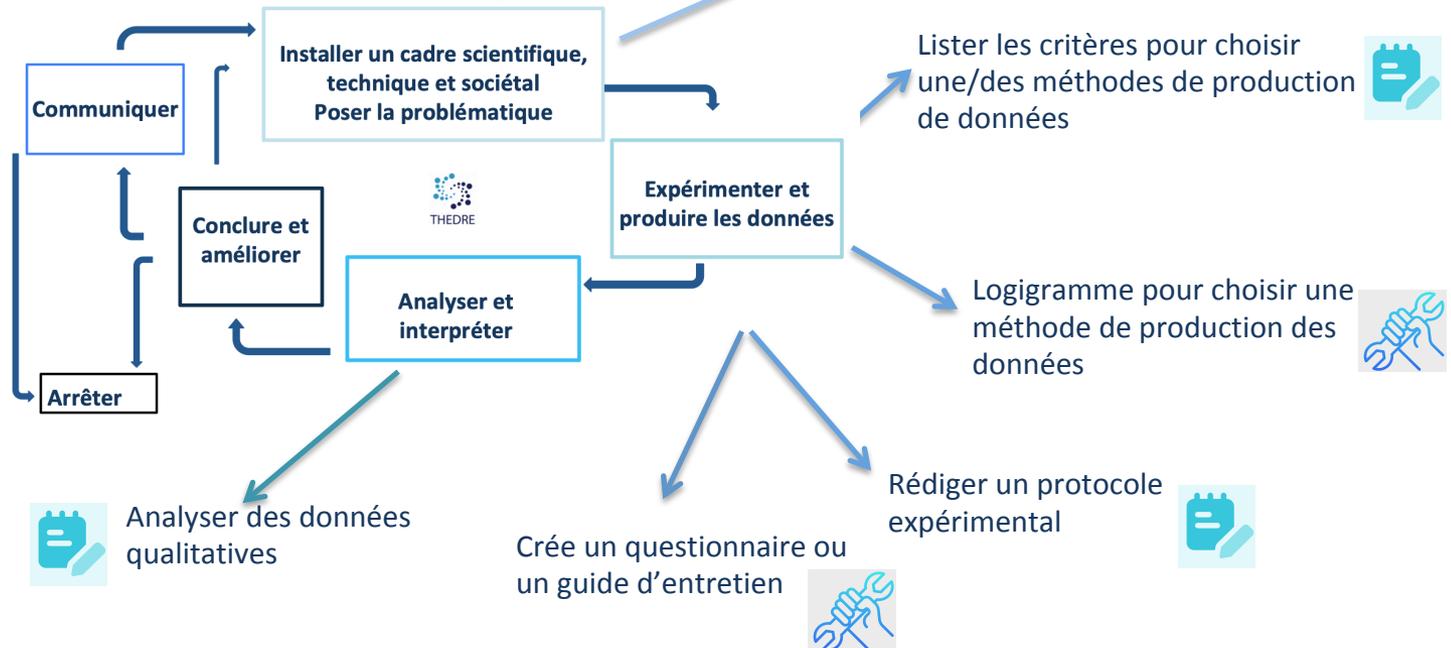
Rédiger un protocole expérimental



Crée un questionnaire ou un guide d'entretien



Analyser des données qualitatives





Discussion ...

Gestion du temps

- Comment faites vous ?





Gestion du temps

- Une vision large
 - Calendrier par semestre ou trimestre
 - Fixer des jalons, des livrables
 - Fixer des objectifs

	An 1	An 2	An 3	An 3,5
Semestre 1				
Semestre 2				



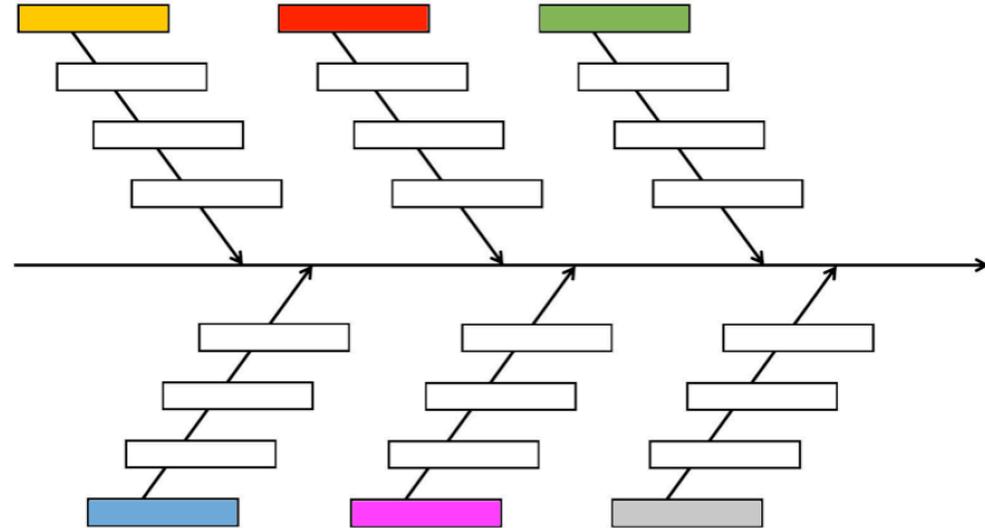
Gestion du temps

- Une vision micro-projets plus facile à gérer
- Quels sont les micro-projets de votre thèse ?



Séquencer son travail

[Ishikawa et Kanban](#)



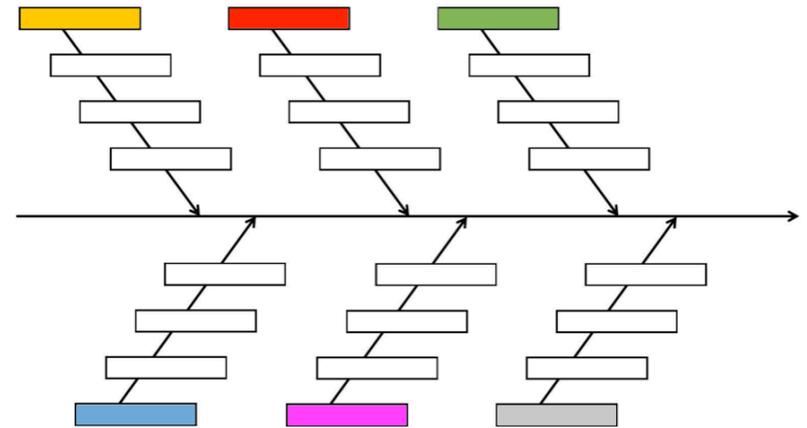
Etapes	A faire	En cours	Fait	Trop tard
Yellow				
Blue				
Red				
Pink				
Green				





Séquencer son travail

[Ishikawa et Kanban](#)



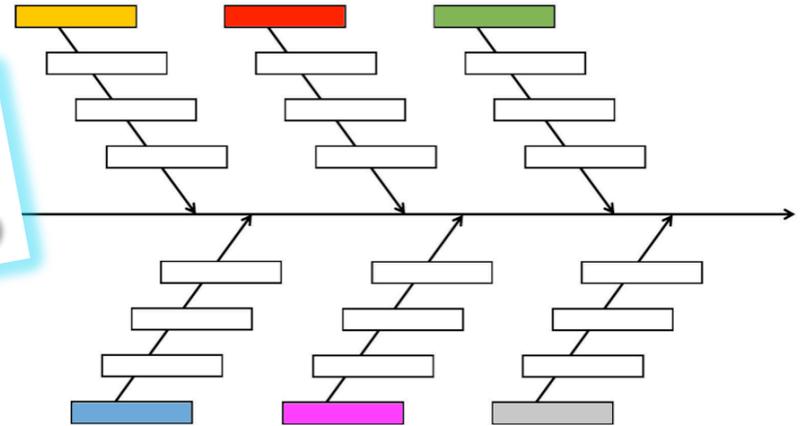
Identifier les micro-projets de votre thèse

Identifier les actions liées à chacun de ces moments



Séquencer son travail

Debriefing



Identifier les moments de votre thèse

Identifier les actions liées à chacun de ces moments





Indicateurs piloter les micro-projets

- Indicateurs qui rend compte votre activité :
 - Ce qui est fait, ce qui est produit
 - Activité : Créer un outil, rédiger le protocole expérimental, Lire la bibliographie,...
 - Production : les données expérimentales, une synthèse de la bibliographie, ...
- Indicateurs d'objectifs
 - Fixés a priori avant le démarrage d 'un projet pour savoir si le projet a atteint son but
 - Exemple : Avoir suivi un dispositif sur 2 ans, nombre d'utilisateurs >1000 étudiants, niveau de performance > 15km/h



Indicateurs piloter les micro-projets

Finalité : Suivre l'activité et le processus de recherche

Propriétés d'un indicateur : **S**pécifique, **M**esurable,
Aceptable, **R**éaliste, **T**emporel

Mesurer lors de l'activité du doctorant et lors des
expérimentations



« Identifier des indicateurs d'activités et d'objectifs »

Ce qu'en disent les doctorants : « La réflexion sur les indicateurs permet de creuser et de se fixer des objectifs plus précis. » ; « Aide à avoir une vision claire des différents points qui entourent un objectif ».

Quand : Ce document doit vous accompagner dès le début de la thèse. Vous pouvez revenir dessus tous les 6 mois ou après chaque expérimentation.

Pourquoi : Il permet de travailler sur la création des indicateurs d'objectifs, d'approfondir les mesures qui seront prises pendant la production des données, d'affiner le protocole expérimental, de suivre l'activité et le processus de recherche.

Apports pour la thèse : Rendre compte du travail effectué pendant la thèse et voir si les objectifs sont atteints.

Préalable : Les indicateurs sont SMART : Spécifique, Mesurable, Acceptable, Réaliste, fixés dans le Temps. Il faut limiter le nombre d'indicateurs, car la difficulté n'est pas de les créer mais de les suivre dans le temps...



Debriefing

« Identifier des indicateurs d'activités et d'objectifs »

Ce qu'en dise les doctorants : « La réflexion sur les indicateurs permet de creuser et de se fixer des objectifs plus précis. » ; « Aide à avoir une vision claire des différents points qui entourent un objectif ».

Quand : Ce document doit vous accompagner dès le début de la thèse. Vous pouvez revenir dessus tous les 6 mois ou après chaque expérimentation.

Pourquoi : Il permet de travailler sur la création des indicateurs d'objectifs, d'approfondir les mesures qui seront prises pendant la production des données, d'affiner le protocole expérimental, de suivre l'activité et le processus de recherche.

Apports pour la thèse : Rendre compte du travail effectué pendant la thèse et voir si les objectifs sont atteints.

Préalable : Les indicateurs sont SMART : Spécifique, Mesurable, Acceptable, Réaliste, fixés dans le Temps. Il faut limiter le nombre d'indicateurs, car la difficulté n'est pas de les créer mais de les suivre dans le temps...





Bien commencer ...

« Pour bien commencer un travail de thèse, construire sa problématique et rédiger son introduction »

Ce qu'en dise les doctorants : « Très intéressant de se questionner sur des aspects précis, cela permet de réfléchir différemment et de voir qu'il y a besoin d'approfondir certains points. » ; « Pratique et peut-être plus facile à présenter (l'avancé de son travail) à son directeur de thèse pour un temps de travail commun. »

Quand : Ce document doit vous accompagner du début à la fin de votre thèse. En début de thèse : il permet de cadrer le problème et construire sa problématique. Au fil du travail : il permet de raffiner la problématique

Pourquoi : Il permet de se poser les bonnes questions en début de thèse et tout au long du travail. Il permet de ne rien oublier lors de la construction de la problématique. C'est un document de cadrage de votre travail de thèse. Il est important de le partager avec vos encadrants de thèse pour faciliter la compréhension sur le problème à traiter.

Apports pour la rédaction de la thèse : Il permet de rédiger l'introduction de thèse et/ou la problématique.

Préalable : En début de thèse vous ne pourrez sans doute pas répondre à toutes ces questions. Mais il est conseillé de revenir régulièrement sur ce document pour affiner la problématique et ainsi rédiger l'introduction de la thèse.



Bien commencer ...



« Pour bien commencer son travail de thèse, construire sa problématique et son introduction »

Debriefing

Ce qu'en dise les docteurs : « Donner sur des aspects précis, cela permet de réfléchir différemment et de voir les choses sous un autre angle. » ; « Pratique et peut-être plus facile à présenter (l'avancé de son travail) à son directeur de thèse pour un temps de travail commun. »

Quand : Ce document doit vous accompagner du début à la fin de votre thèse. En début de thèse : il permet de cadrer le problème et construire sa problématique. Au fil du travail : il permet de raffiner la problématique

Pourquoi : Il permet de se poser les bonnes questions en début de thèse et tout au long du travail. Il permet de ne rien oublier lors de la construction de la problématique. C'est un document de cadrage de votre travail de thèse. Il est important de le partager avec vos encadrants de thèse pour faciliter la compréhension sur le problème à traiter.

Apports pour la rédaction de la thèse : Il permet de rédiger l'introduction de thèse et/ou la problématique.

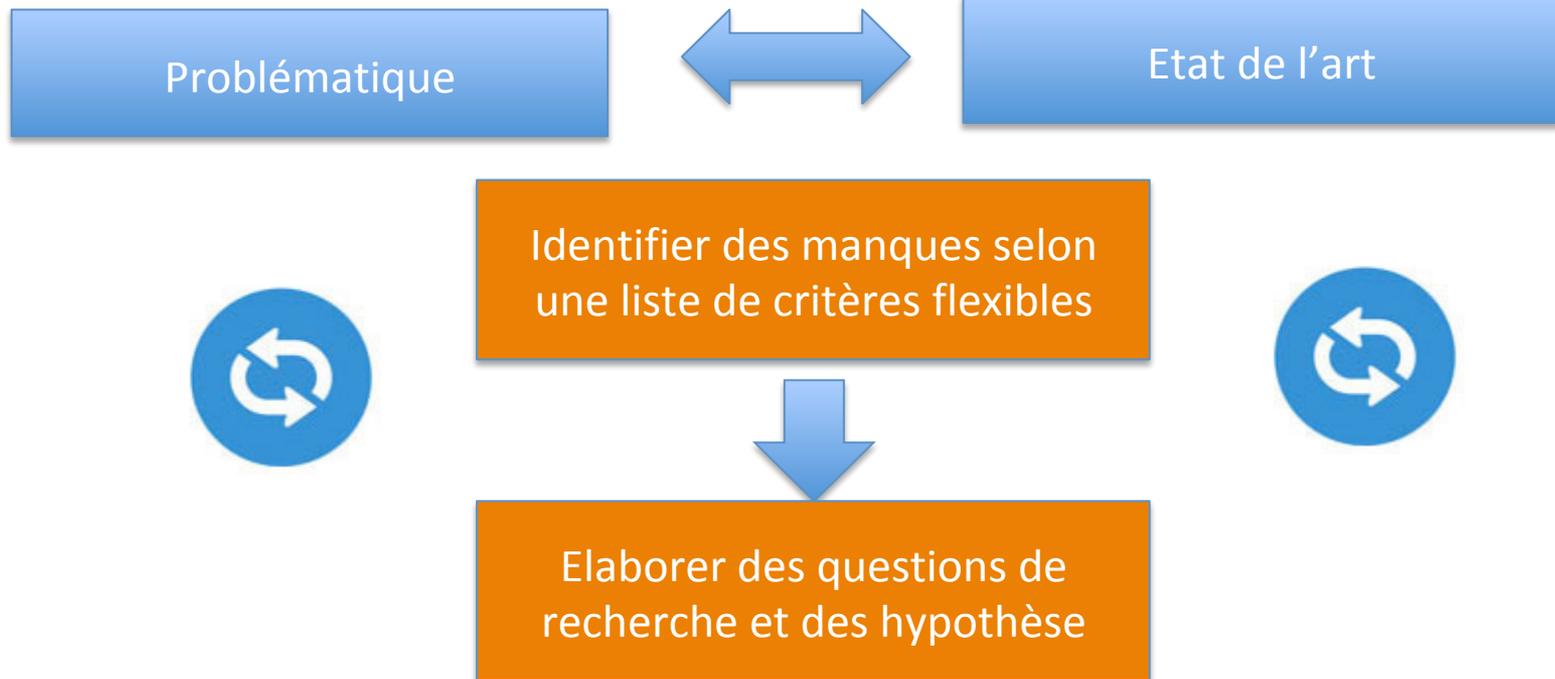
Préalable : En début de thèse vous ne pourrez sans doute pas répondre à toutes ces questions. Mais il est conseillé de revenir régulièrement sur ce document pour affiner la problématique et ainsi rédiger l'introduction de la thèse.





Etat de l'art

Pourquoi ?





Etat de l'art Comment?



	Article 1	Article 2	Article 3	Article 4
Critère initial n°1		x		
Critère initial n°2	x		x	x
Critère initial n°3			x	
Critère initial n°4	x			



Etat de l'art Comment?



	Article 1	Article 2	Article 3	Article 4
Critère initial n°1		x		
Critère initial n°2	x		x	x
Critère initial n°3			x	
Critère initial n°4	x			
Nouveau critère 1		x	x	x
Nouveau critère 2		x		



Etat de l'art Comment?



	Article 1	Article 2	Article 3	Article 4
Critère initial n°1		x		
Critère initial n°2	x		x	x
Critère initial n°3			x	
Critère initial n°4	x			
Nouveau critère 1		x	x	x
Nouveau critère 2		x		



- Guide pour conduire un état de l'art
- Carte mentale



Produire des données

- Quelles sont les méthodes de production de données que vous connaissez ?



Expérimenter

Que mettre dans un protocole expérimental ?



Expérimenter

« Rédiger un protocole expérimental »

Laissons parler les doctorants : « *C'est intéressant parce qu'il permet de se questionner sur des points plus avancés.* » ; « *Bon exercice à faire en début de thèse.* » ; « *Permet de mettre le doigt sur les manques d'informations personnelles nécessaires à la constitution d'un protocole de récolte de données* »

Quand : Ce document doit vous accompagner dès que les outils à tester sont créés, quand les méthodes de production de données ont été choisies.

Pourquoi : Il permet de rédiger le protocole expérimental pour produire des données et spécifier les mesures et les méthodes de production et d'analyse des données.

Apports pour la rédaction : Il est à utiliser pour rédiger les parties méthodologiques de la thèse.



Expérimenter

Debriefing

« expérimental »

Laissons parler les doctorants : « C'est intéressant parce qu'il permet de se questionner sur des points plus avancés. » ;
« Bon exercice à faire en début de thèse. » ; « Permet de mettre le doigt sur les manques d'informations personnelles
nécessaires à la constitution d'un protocole de récolte de données »

Quand : Ce document doit vous accompagner dès que les outils à tester sont créés, quand les méthodes de
production de données ont été choisies.

Pourquoi : Il permet de rédiger le protocole expérimental pour produire des données et spécifier les mesures et les
méthodes de production et d'analyse des données.

Apports pour la rédaction : Il est à utiliser pour rédiger les parties méthodologiques de la thèse.



Bilan expérimental



Comment faire un bilan expérimental ?



Bilan expérimental

« Faire le bilan des expérimentations »

Quand : Ce document doit être rédigé à la fin d'une expérimentation avant de commencer le traitement des données.

Pourquoi : Il permet de faire le bilan de ce qui s'est produit pendant l'expérimentation, de relever les points forts et les lacunes de l'expérimentation. Rédiger juste après l'expérimentation, il permet de noter les impressions sur l'expérimentation.

Apports pour le traitement des données : La rédaction de ce document facilite le traitement des données et l'interprétation des résultats.

Apports pour la rédaction de la thèse : Il servira à la rédaction de la partie expérimentation du manuscrit et il indiquera les limites de validité des données recueillies.



Bilan expérimental



« Faire le bilan des expérimentations »

Debriefing

Quand : Ce document est rédigé avant de commencer le traitement des données.

Pourquoi : Il permet de noter ce qui s'est produit pendant l'expérimentation, de relever les points forts et les lacunes de l'expérimentation. Rédiger juste après l'expérimentation, il permet de noter les impressions sur l'expérimentation.

Apports pour le traitement des données : La rédaction de ce document facilite le traitement des données et l'interprétation des résultats.

Apports pour la rédaction de la thèse : Il servira à la rédaction de la partie expérimentation du manuscrit et il indiquera les limites de validité des données recueillies.





Analyses des Données Quantitatives

Quelles sont les étapes pour faire une analyse des données ?



Validation des données

Travaux de Laure Berti Equille

Comme le représente la Figure 1, on peut classer la plupart des travaux abordant la problématique de la qualité des données selon quatre grands types d'approches complémentaires.

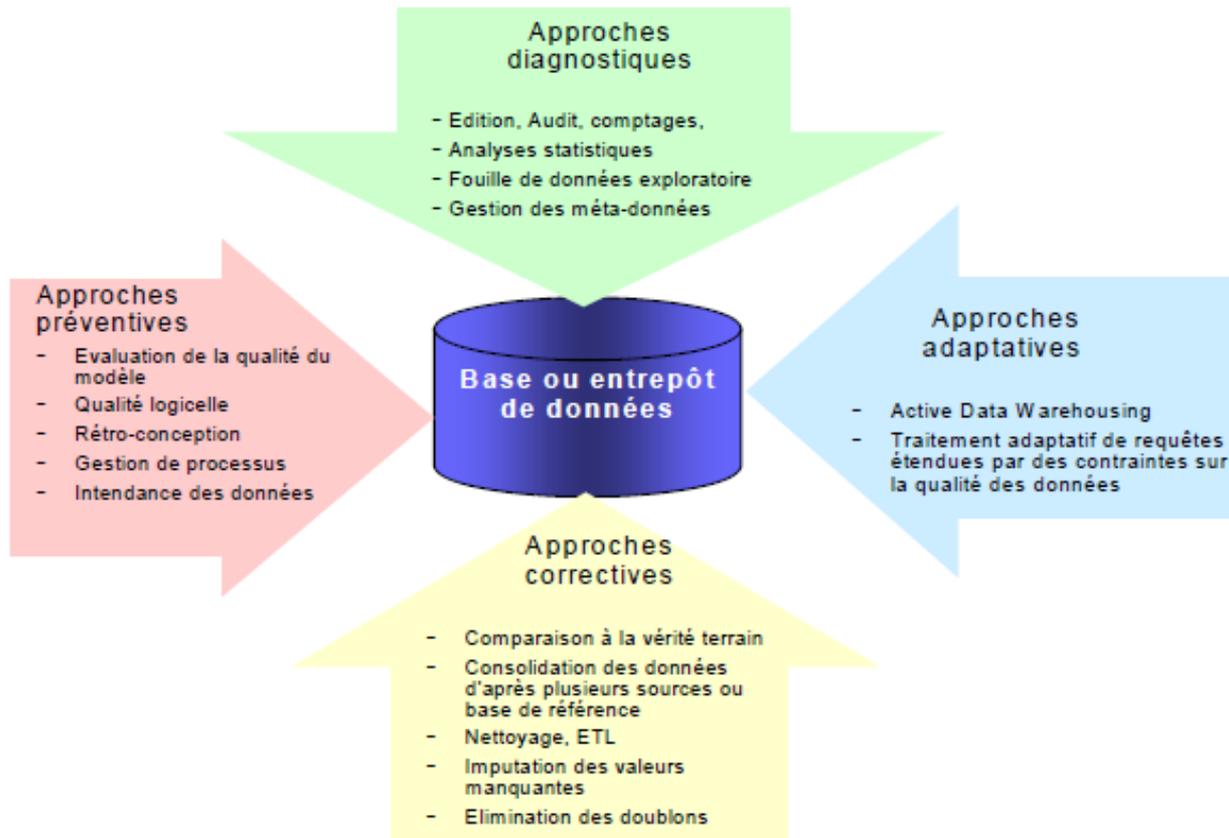


Figure 1 - Panorama des approches pour l'évaluation et le contrôle de la qualité des données



Indicateurs de qualité des données

- **Pertinence** : capacité des données répondre aux besoins actuels et futurs des utilisateurs.
- **Exactitude/Justesse** : mesure de la conformité des données par rapport à la réalité. (p.ex., la taille des individus ne peut pas être supérieur à 2,50 m ni inférieur 0,40 cm).
- **Précision temporelle** : exactitude des données par rapport à l'instant qu'elles sont censées représenter. Le chercheur a besoin d'avoir des données qui décrivent une situation telle qu'elle est ou était à un moment précis. (p.ex., les bilans pour une entreprise sont enregistrés avec l'année de référence).
- **Accessibilité** : la facilité de localisation et d'accès aux données et aux métadonnées.
- **Facilité d'interprétation** : facilité de compréhension des données, de leur analyse et de leur usage. Les données doivent être bien documentées pour être comprises sans ambiguïté.
- **Unicité** : garantie qu'une entité du monde réel est représentée par un seul et unique objet, il s'agit de contrôler la présence des doublons.



Indicateurs de qualité des données

- **Cohérence** : absence d'informations conflictuelles. (p.ex., l'âge des enfants doit être inférieurs à celui de leurs parents).
- **Conformité à une norme** : respect d'une norme standardisée ou d'une convention de nommage (p.ex., la profession de la personne est codée selon la norme INSEE : PCS en 8 catégories).
- **Complétude** : Ce critère est utilisé dans les approches préventives. Car il s'agit de contrôler si les objets nécessaires à la production des données sont présents dans le modèle de données. La complétude se juge en fonction selon 4 critères : entités, attributs, relations et occurrences. (p.ex., pour les entités, une base de données des clients est incomplète s'il manque l'adresse de facturation, p.ex., pour les relations, une personne peut aller dans plusieurs salles de cinéma, le modèle doit comporter une relation « voir des films » liant les entités « personne » aux entités « salles de cinéma »).
- **Consistance** : Quand une entité est recopiée, il y a consistance si on retrouve les mêmes valeurs d'attributs dans toutes les bases.
- ambiguïté.



Qqs références

- Berti-Equille, L. (2007). Quality awareness for managing and mining data [Habilitation à diriger des recherches, University Rennes 1]. http://www.researchgate.net/publication/251573080_L._Berti-quille_Quality_Awareness_for_Data_Managing_and_Mining_Habilitation__Diriger_des_Recherches_Universit_de_Rennes_1_Juin_2007/file/5046351f12d600266b.pdf
- Berti-Equille, L. (2012). La qualité et la gouvernance des données: Au service de la performance des entreprises. Hermes Science Publications.
- Di Ruocco, N., Scheiwiler, jean-M., & Sotnykova, A. (2012). La qualité des données: Concepts de base et techniques d'amélioration. In La qualité et la gouvernance des données (Hermes, pp. 25–55). Lavoisier.





Analyses des Données Quantitatives

« **Bien analyser des données quantitatives et documenter ce travail** »

Laissons parler les doctorants : « *Je ne sais pas comment faire pour analyser mes données* », « *ma p valeur est-elle bonne ?* »

Quand : Ce document doit vous accompagner lorsque que vous allez commencer le traitement de vos données.

Pourquoi : Il permet de documenter la façon dont vous allez valider, pré-traiter et traiter vos données.

Apports pour la rédaction de la thèse : Il permet de rédiger la partie traitement des données et résultats. Il vous permet aussi de rédiger la partie résultats d'un article de recherche.

Des plateformes de type Jupyter peut vous aider à documenter ce travail.

Objectifs du traitement

- Pouvez-vous rappeler les objectifs de votre expérimentation ?
- A quelles questions/hypothèses votre traitement de données doit-il répondre ?



Analyses des Données Quantitatives

« Bien analyser des données quantitatives et documenter ce travail »

Débriefing

Laissez-vous guider par la question « comment faire pour analyser mes données », « ma p valeur est-elle bonne ».

Quand ? Le débriefing doit vous accompagner lorsque que vous allez commencer le traitement de vos données.

Pourquoi : Il permet de documenter la façon dont vous allez valider, pré-traiter et traiter vos données.

Apports pour la rédaction de la thèse : Il permet de rédiger la partie traitement des données et résultats. Il vous permet aussi de rédiger la partie résultats d'un article de recherche.

Des plateformes de type Jupyter peut vous aider à documenter ce travail.

Objectifs du traitement

- Pouvez-vous rappeler les objectifs de votre expérimentation ?
- A quelles questions/hypothèses votre traitement de données doit-il répondre ?



Créer un outil de mesures en SHS

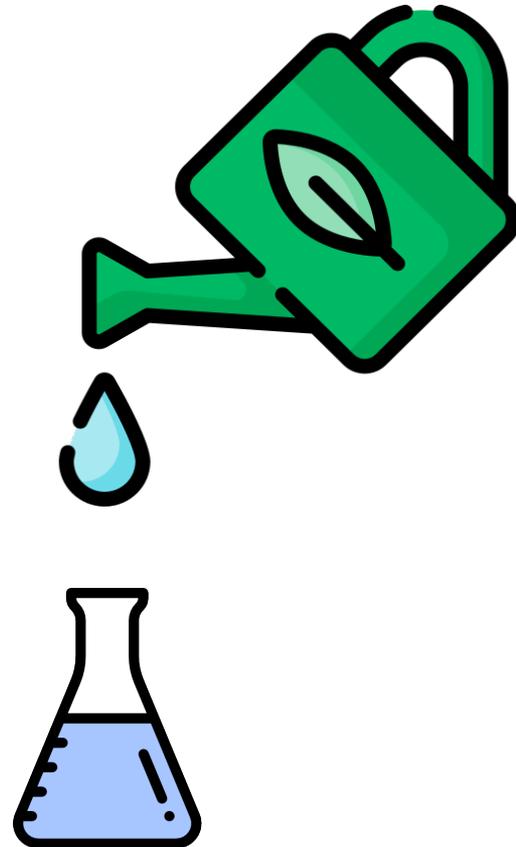
Entretien

Et/Ou

Questionnaires

Et/Ou

Traces



Entretien

Etudier les pratiques des enseignants avec des pédagogies actives et des outils numériques

Enseignement : L1, en Biologie, UE sciences expérimentales. Responsable d'UE veut introduire une nouvelle plateforme existante

Le responsable d'UE est intéressé par les méthodes Student Centered Learning (Freeman 2014)



Entretien

Lister les mesures à prendre

- Pratiques des pédagogies actives
- Avantages, inconvénients de ces pédagogies
- Usage du numérique dans la classe
- Avantages, inconvénients du numérique
- Perception *a priori* de la nouvelle plateforme
- Perception sur les pédagogies actives Student Centered Learning (autonomie, collaboration, organisation, ...)

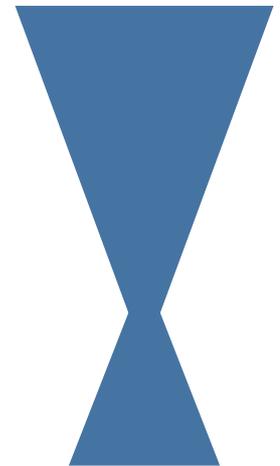
Organiser et ordonner les mesures par grands thèmes

Pédagogies actives

Usage du Numérique

Méthode Student Centered Learning

La nouvelle plateforme



Guide d'entretien 1/4

C'est un document rédigé, qui est testé

Les questions s'enchaînent, pour arriver au point essentiel de l'entretien

Faire une introduction, sur la raison de l'entretien et comment il va se dérouler

- Il faut présenter au sujet pourquoi nous lui avons demandé de participer à cet entretien
- En faisant attention à ne pas lui donner trop d'informations qu'il cherchera à valider pour vous faire plaisir pendant l'entretien
- lui indiquer les étapes de l'entretien
- Lui dire que c'est votre proposition qui est évaluée et pas lui
- Lui dire d'exprimer tout ce qu'il a envie de dire par rapport à la proposition

Guide d'entretien 2/4

- **Du spontané au suggéré**
 - Laisser libre cours au discours de la personnes
 - Si elle ne s'exprime pas sur un sujet alors lui poser la question
 - Lors de la rédaction de la grille : **envisager tous les cas de figures.**
 - *Si le sujet a parlé de la méthode Student Centred Learning (peu probable), s'il parle de laisser autonome les étudiants, ...*
 - *Si elle n'en a pas parlé*
- Le guide doit contenir les **remarques pour l'enquêteur**
 - *ENQUETEUR : Noter le déroulement du cours*
 - *ENQUETEUR : Noter les outils numériques cités*
- Le guide se termine en demandant les caractéristiques des participants
 - Année de naissance, vous êtes un homme/ une femme, votre profession, niveau d'étude, domaine d'activités, ancienneté dans la profession , votre type d'habitat, nombre d'enfants ... et toutes informations utiles pour répondre à vos questions.

Guide d'entretien 3/4

Les domaines qui sont abordés sont parfois complexes pour cela deux possibilités non exclusives :

Proposer des constats pour lancer l'entretien

Selon la directive, XYZ de 2018, Les pédagogies actives sont de plus en plus présentes dans l'enseignement supérieur. Elles sont souvent utilisées dans des contextes de spécialistes.

Proposer des définitions sur des termes qui ne doivent pas être ambigus

*Aujourd'hui, on entend beaucoup parler de **réseaux sociaux**. Dans notre étude, il s'agit des réseaux de type facebook où chacun est libre de déposer des informations qui le concerne et de les partager avec d'autres personnes. L'utilisateur de facebook peut également créer des groupes de personnes auxquels il réservera certains types d'informations .*

Guide d'entretien 4/4

- **Rédaction des questions**
 - Les questions doivent permettre des **réponses larges et ouvertes** :
 - *Avez vous déjà mis en place des pédagogies actives ?*
 - *Dans quelles circonstances avez vous mis en place des pédagogues actives ? (oui)*
 - **Ne jamais suggérer** de réponses dans les questions
 - *Pensez-vous que ce dispositif soit utile ? (non)*
 - *Que pensez-vous de ce dispositif ? (oui)*

Entretien : Passation

- **Faire un pilote**, un test avec une personne pour identifier si les questions sont comprises et si l'ordre des questions est suffisamment fluide.
- **Connaître** son guide d'entretien, à partir de 3 ou 4 passations, on le connaît
- Prévoir un **enregistrement** des entretiens, demander l'accord au moins oral aux personnes
- **Prendre des notes et laisser s'installer des silences**
- Lors de la passation de l'entretien, l'interviewé peut parfois aborder les différents éléments prévus par la grille sans forcément respecter l'ordre que vous avez prévu
- Peu importe : l'entretien est semi-directif ce sera l'enquêteur de conduire l'entretien sans pour autant être dirigiste.
- Laisser le sujet exprimer ses idées même si elles ne sont pas dans l'ordre de la grille. L'enquêteur doit veiller à ce que tous les points soient abordés.
- **Ne pas se justifier** en cours d'entretien, attendre la fin de l'entretien pour expliquer à

Questionnaire 1/2

- **Généralités**

- Doit comporter une introduction,
- La structuration reste la même du plus global au plus précis,
- Les modalités des réponses sont définies, à l'exception des questions ouvertes,
- Les variables sociodémographiques se situent à la fin.

- **Types de questions**

- **fermées** (close-ended question) : les réponses aux questions sont fixées, il n'est pas possible de rajouter des modalités de réponses
- **numériques**
 - Combien de voitures possédez-vous ?
- **semi-ouvertes** (semi-open question): questions pré codées, avec l'option "autre" qui permet d'ajouter des items de réponse.
 - Quels sont vos sports favoris :
 - Course à pied
 - Ski
 - Natation
 - Autres précisez
- **ouvertes (open-ended question)** : l'enquêté répond librement, pas de propositions de réponses a priori.

Questionnaire 2/2

- Réponse unique sur une échelle de réponse (scale) :
 - Echelle de Likert (Likert scale) La personne interrogée exprime son degré d'accord ou de désaccord vis-à-vis d'une affirmation. L'échelle contient en général cinq ou sept choix de réponse qui permettent de nuancer le degré d'accord. Pas du tout d'accord, Pas d'accord, Ni en désaccord ni d'accord, D'accord, Tout à fait d'accord. Pour les échelles impaires, le niveau central permet de n'exprimer aucun avis, tandis que les échelles paires (par exemple à quatre modalités) sont dites « à choix forcé ». (5)
- Réponse à choix multiples (multiple-choice questions) :
 - Choix de plusieurs modalités de réponse : (exemple j'utilise mon portable pour 1- Téléphoner 2 - Envoyer des SMS 3- Pour tenir mon agenda, 4-Pour me réveiller, 5- Autres)
 - À opposer aux questions à choix unique, pas de possibilité de rajouter une modalité de réponse
- Réponse avec une demande de classement :
 - les réponses sont ordonnées selon une préférence ou un ordre d'importance. :
 - *Pour le produit XYZ vous classerez les 3 caractéristiques suivantes en partant de ce vous considérez comme son point le plus fort vers son point le plus faible :*
 - *Facilité d'utilisation*
 - *Fiabilité d'utilisation*
 - *Son prix*



Traces

- Un guide de production des traces
 - Liste des variables utiles pour l'analyse
 - Temporelle : Time stamp, **et** no d'ordre
 - Utilisateurs : Code enseignant, apprenant, ...
 - Actions avec l'outil : nom de l'action, ...
 - Domaine de validité de ces variables
 - Age compris entre 18 et 25 ans
 - Heure de connexion entre 14h et 18h



- Rédiger quelques questions pour un guide d'entretien sur votre cas ?
- Rédiger quelques questions pour un questionnaire ?

Tracer l'activité des focus groups avec des activités des utilisateurs

No d'étape : Indiquer un numéro d'étape

Objectif de l'étape : Objectifs généraux de l'étape

Hypothèses et questions en lien avec cette étape : Indiquer les questions ou les hypothèses expérimentales abordées dans cette étape. Ces questions et hypothèses correspondent à celles posées dans le protocole expérimental.

Activités réalisées : Lister activités faites pendant cette étape. (p.ex., présentation du modèle, tests utilisateur, lecture du dictionnaire)

Responsables de la passation : Nom de la personne en charge de faire la passation de cette étape

Acteurs concernés: Chercheur/Méthodologue/Développeur/Utilisateur/autres préciser

Composant(s) activable(s) : Indiquer le(s) composant(s) activable(s) concernés par l'étape.

Format du composant : Indiquer l'état d'avancement de(s) composants : dynamique/statique/ non manipulable

Responsables du (des) composant(s) : Noms des personnes en charge de faire le(s) composant(s)

Matériel expérimental : Liste des matériels expérimentaux nécessaires à cette étape (p.ex., guide d'entretien, d'annotation). Indiquer le rôle de ces documents (présentation, recueil).

Responsable du matériel expérimental : Noms des personnes en charge de faire le(s) matériels expérimentaux

Durée (mn) prévue Indiquer une durée approximative de l'activité

Heure début L'heure de début de l'étape

Heure fin L'heure de fin de l'étape

Durée cumulée prévue (hh:mm): Cumul de la durée des activités



Processus d'évaluation longitudinal, (Mandran, et al 2019, EIAH)

- Les mesures de la littérature
 - Utilisabilité (SUS Brooke)
 - Student centered Learning (Hannafin 2012, Freeman 2014)
- Des spécifiques
 - Satisfaction, attentes, pratiques
- Des outils de mesures

Mesures quantitatives	Mesures qualitatives
Questionnaires baromètres (QB) Questionnaires de satisfaction (QS) Traces d'activité	Entretiens semi-directifs Analyse des missions Débriefings des équipes pédagogiques

- Le processus

	Avant utilisation	Année 1	Entre années 1 & 2	Année 2	Entre années 2 & 3	Année 3	Après utilisation
Responsables d'UE	QB0 Entretien initial		QB1 Analyse missions 1		QB2 Analyse missions 2		QB3 Analyse missions 3 Entretien final
Autres enseignants	QB0		QB1 Débriefing 1		QB2 Débriefing 2		QB3 Débriefing 3
Etudiants		Traces 1	QS1	Traces 2	QS2	Traces 3	QS3





Après la production, l'analyse et la rédaction





Analyse Qualitative

Entretien : Analyse thématique simplifiée

Méthode en Sociologie :

Exemple

Retranscrire les entretiens (1h => 6h)

A partir de la problématique et de la grille d'entretien, créer une grille thématique,

A partir de la **retranscription des entretiens**, identifier les thèmes qui correspondent à la problématique et aussi les thèmes qui émergent

Les thèmes sont larges, il convient de créer des sous-thèmes

Chacun des sous-thèmes est illustré par des verbatims (extraits du discours associé au code du sujet ayant prononcé de discours)

Synthèse pour chaque thème

Thème « Utilisation du numérique dans l'enseignement »

Thème nouveau « Utilisation que du numérique »

Sous- thèmes de « utilisation du numérique »

« utilisation volontaire »

« utilisation contrainte par le responsable »

Verbatim du sous- thème « utilisation contrainte par le responsable »

« c'est le responsable de cette UE cette année qui a choisit de mettre cela en place ans les cours, je ne suis pas certain que ce soit utile et que cela change quelque chose »(S1) »



Analyse Qualitative

Schéma

Créer une grille d'analyse pour savoir ce qui doit être repéré par rapport à la problématique

Photographier chacune des productions

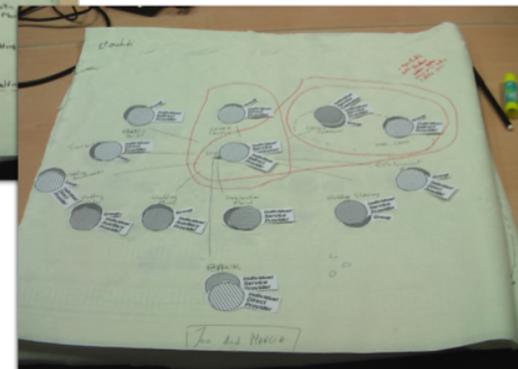
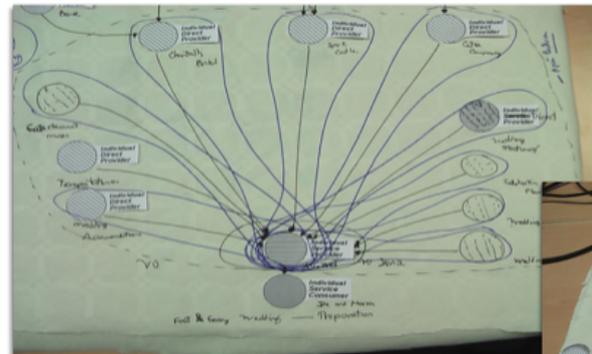
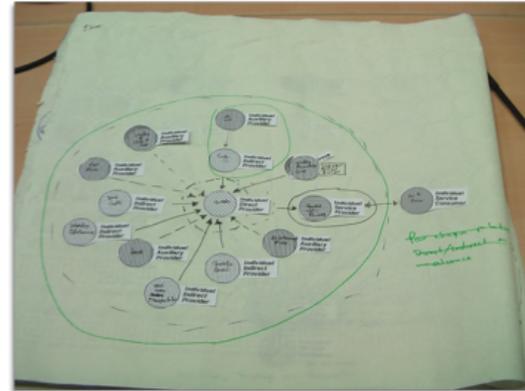
Repérer les ressemblances/les différences entre les représentations

Repérer ce qui confirme les objectifs de l'étude

Repérer les apports de la représentation des sujets par rapport à la proposition

Visionner et écouter les films

Conduire une « analyse thématique simplifiée »





Analyse Qualitative

Annotation Audio ou Video

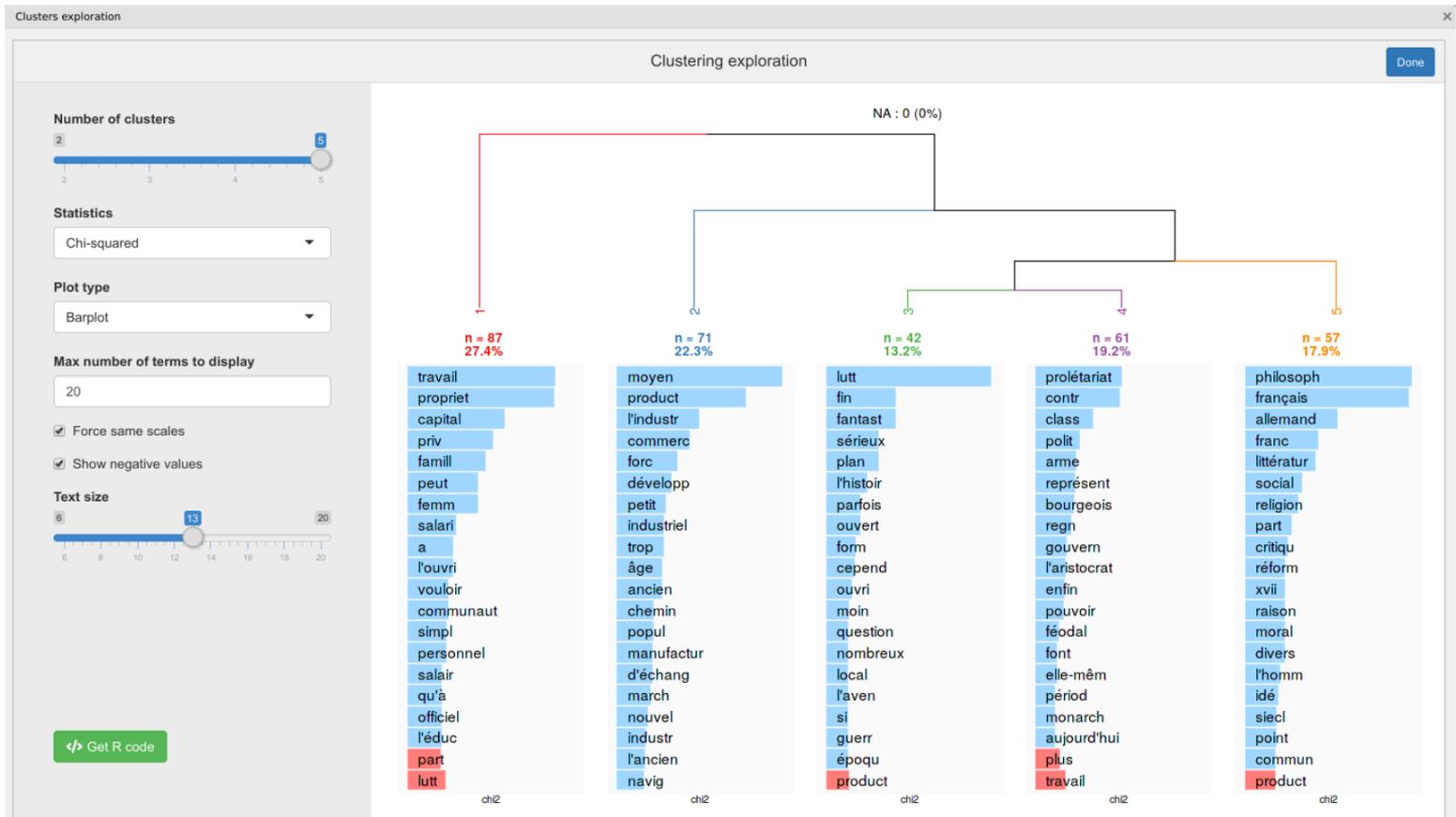
- Créer une grille d'analyse pour savoir ce qui doit être repéré par rapport à la problématique.
- Utiliser des logiciels d'annotations de video
- Analyser les annotations
 - Statistiques descriptives
 - Extraction de verbatims
- Faire une synthèse

The screenshot displays the ELAN 3.9.1 software interface. The main window shows a video of two people sitting and talking. The interface includes a menu bar (File, Edit, Annotation, Tier, Type, Search, View, Options, Window, Help) and a toolbar with various playback and editing controls. The 'Audio Recognizer' tab is active, showing parameters for 'Tag vowels (volume peaks of voiced timespans)'. The 'Parameters' section includes sliders for 'Pitch ceiling [Hz]' (set to 604.8) and 'Intensity change [dB] to start/end a peak (0.5 - 5.0)' (set to 2.0). The 'Progress' bar indicates the recognizer is 'Ready'. Below the video, there are several tracks: a green waveform for intensity, a black waveform for amplitude, and a timeline with various annotations. The timeline includes a 'Clause Transcri' track with text segments like 'and so he climbs up a tree and he starts with the ladder', a 'Motion' track with 'motion' and 'non-motion' labels, and a 'Gesture #' track with 'gesture 1' through 'gesture 9' labels. The bottom left corner shows a legend for the annotation tracks.



Analyse textuelle

Alceste
RThemis
Rainette





Analyse Qualitative

Points forts et Points faibles

Outil : Grille qui liste les éléments relatifs à une proposition et qui à deux colonnes une pour noter les points forts et l'autre les points faibles

Exemple :

Table Number :

First and last name :

<i>List the strengths and weaknesses on the following</i>	Strengths	Weaknesses
the work performed with the method	+	-
the collaborative work enabled with the method	+	-
The creativity enabled with the method	+	-
the serious gaming of the method	+	-
The software compared to card game	+	-



Analyse Qualitative

Points forts et Points faibles

Méthode :
Saisir l'ensemble des réponses : Ligne=participants, Colonne = réponse aux questions, 2 colonnes par éléments

Experimentation menée en 2011 lors de l'école d'été sur les GBL. LIG/METAH Pernin et al							
				The collaborative work enabled with the method		The method ability to give new ideas	
				What you like +	What you don't like -	What you like +	What you don't like -
saisie	su	tab	ses				
CM	1	1	2	It's easy to work collaboratively	Difficult to agree in the selected cards	very intuitive to create many different ideas	
CM	2	1	2	immediate reviewing	maybe it can be difficult to make the game with people with different backgrounds	the fact of having one expert facing newbies on the subject	
CM	3	1	2	The cards are good support to communicate with other members of the team. Rapid modification / iteration	If you move the table everything moves and you have to replace the cards / jewels	the existing game principles and cards help us to set up some foundations	can also avoid some possible activities or scenarios because of the formalism. If / esle ? Random ?
CM	1	2	2	great for triggering discussion. The terminology might need to be tightened. Capitalize might not work that well.	Potential for one person to dominate the interaction physically (but not completely)	it was very helpful. I have a lot of new ideas I may follow up. Perhaps more creative game scenarios in activity 1 would be useful for inspiration.	need more time afterwards for reflection + exploring ideas that arise
CM	2	2	2	we shared many different ideas	we should take some roles so we could be more organized	we produced a lot of ideas	
CM	3	2	2	cards facilitate greatly the collaborative approach.		this method brings new ideas because students start on groupware gaming activities and their synchronisation and communication issues.	the method should be extendible ; introduce profiles.
CM	4	2	2	I like combination of different opinions of participants		i like to see differents point of view	
CM	1	2	3	échanges nombreux ; matériel (forte flexibilité dans la manipulation des différents éléments) a aidé à la collaboration	auto-régulation => en cas de désaccord, pas de moyen prévu pour aider à trouver des compromis	grandes variétés de choix; Prévoit différentes méthodes de travail, etc.	cadre tout de même très normatif trop contraignant, il est difficile de se conformer aux cartes dans la phase de conception globale
CM	2	2	3	la répartition des rôles ; le retour collectif	pas assez détaillé dans la répartition des rôles	l'encadrement proposé	
CM	3	2	3	visualisation de la progression, des itérations	n'aide pas forcément à la collaboration quand les mécaniques de jeu sont machées dès la début	on essaie de partir des ressorts de jeu	donner des exemples de jeu pour faciliter les idées





Analyse Qualitative

Points forts et Points faibles

The collaborative work enabled with the method

Les points positifs

3 cases vides (dont 2 vides aussi côté points négatifs)

Méthode :
Faire une « analyse
thématique simplifiée » et une
synthèse

1. Concernant la méthode		
La méthode facilite et stimule la collaboration des formateurs	great for triggering discussion The method promotes collaboration. Everyone participating. Good to develop team spirit Easy, nice, good way	14
Elle permet de tous participer, de partager des idées différentes et d'être créatifs collectivement	very good- helped to see each participants point of view and thus come up with a collective work Effective debating, compromising, selection and implementation We all participate in the creation process by bringing ideas and say what we think is not good at other ideas emergence of the solution by speaking and interacting confront, exchange, share ideas, créativité, débat, opinion, brainstorm, open discussion friendly	17



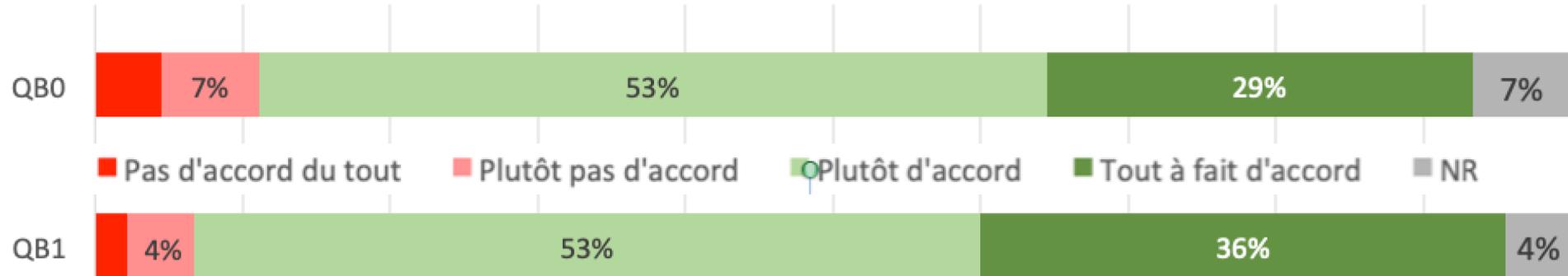
Analyse Quantitative

- Cycle de vie des données => importance du pré-traitement
- Echantillon
- Type de variables => quali/quant/ordinales
- Rôles des variables => explicatives et à expliquer
- Règles de décision sur le choix de la méthode



Analyse Quantitative

"Je pense que LabNbook sera (est) utile pour mes pratiques d'enseignement"

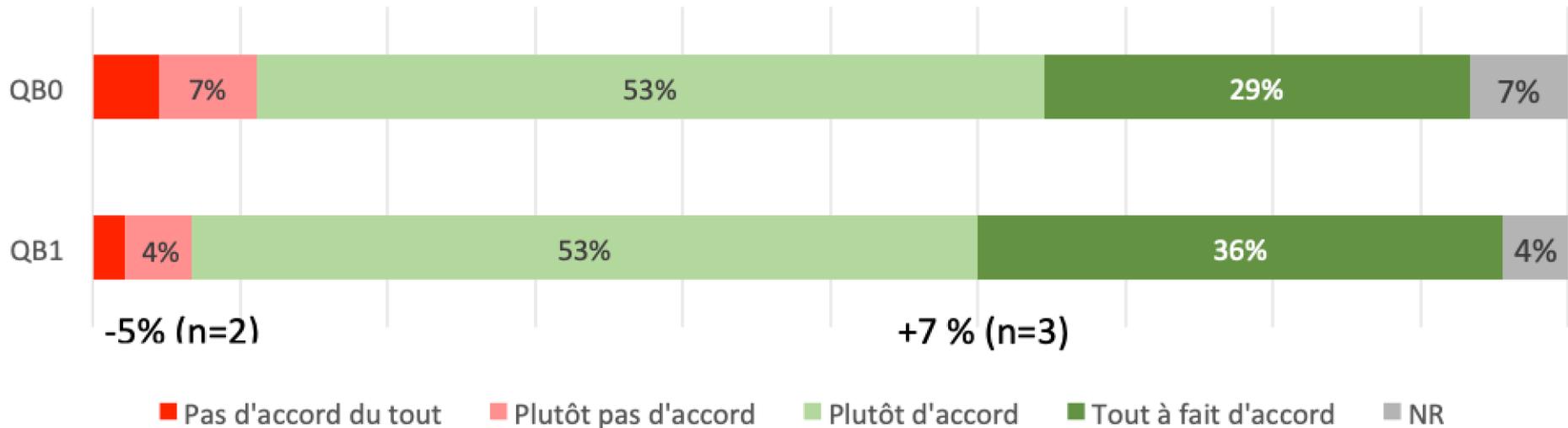


Votre Analyse ??



Analyse Quantitative

"Je pense que LabNbook sera (est) utile pour mes pratiques d'enseignement"



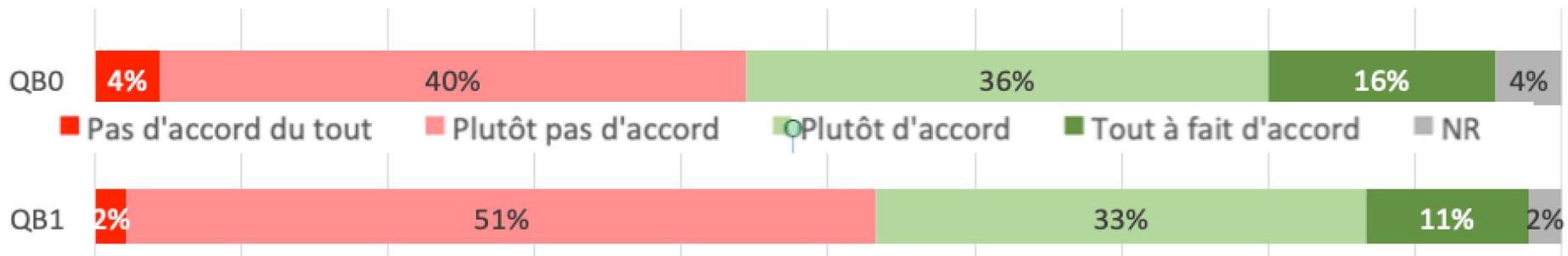
- Evolution positive concernant l'utilité de LabNbook : + 7% d'enseignants sont « *tout à fait d'accord* » avec la proposition

Les chiffres sont affichés en % pour une meilleure lisibilité, mais l'effectif est de 45 personnes, c'est pourquoi nous indiquons les effectifs à coté des %



Analyse Quantitative

« L'utilisation de LabNbook me demandera (me demande) un investissement important en terme de formation »

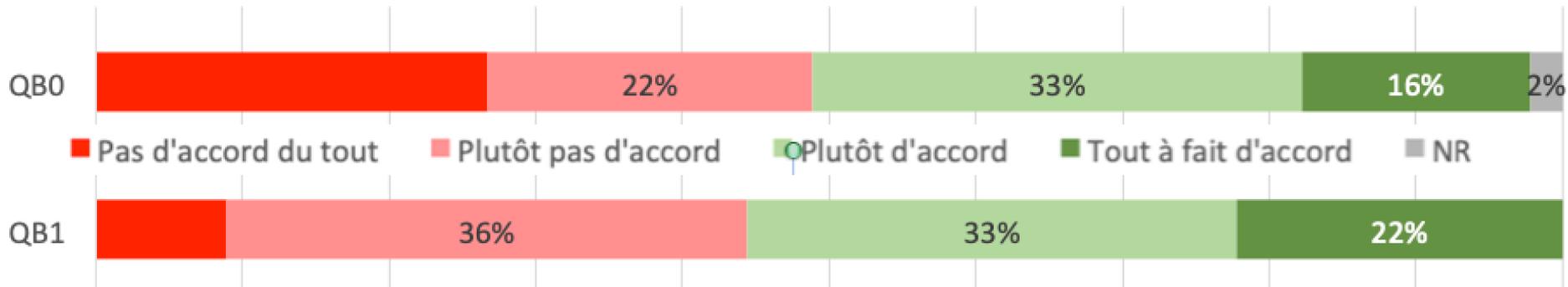


Votre Analyse ??



Analyse Quantitative

« Pour cette formation, je permets (j'ai permis) aux étudiants de travailler de manière itérative »



Votre Analyse ??



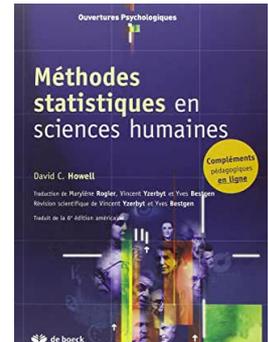
Et en bonus, quelques de mots de stats ...

Méthodes statistiques en sciences humaines, Howell DC : 6^{ème} édition, 2015. (traduit de l'anglais) Edition DeBoeck.

*De l'échantillonnage, **Statistiques descriptives, Visualisations, Corrélations, Statistiques Inférentielles, Modèle linéaire généralisé (ANOVA, régression, ...)**,*

Analyse factorielles simples et multiples. B.Escoffier & J. Pages. 6^{ème} édition. 2016. Edition Dunod.

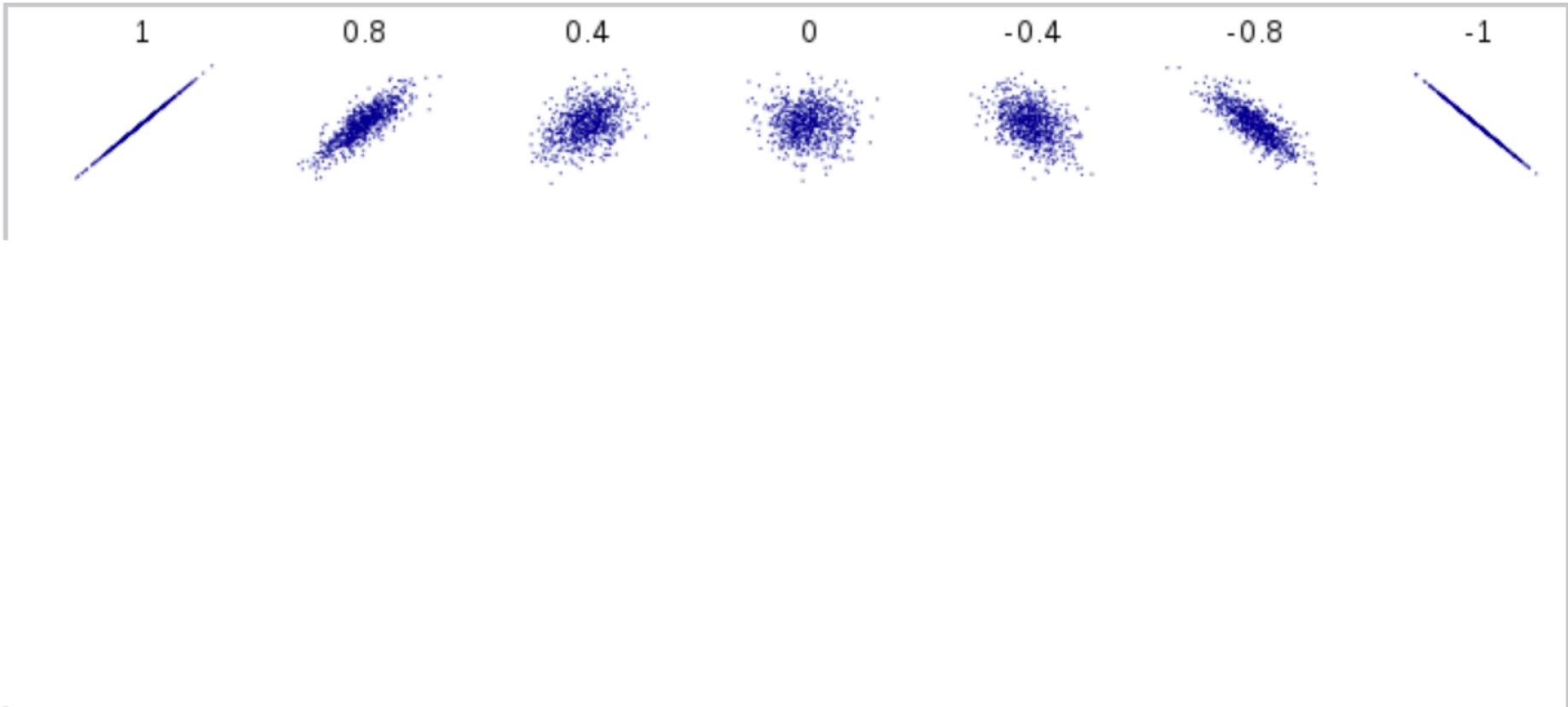
Statistiques exploratoires multidimensionnelles. L.Lebart, A.Morineau, B.Piron. Edition Dunod.



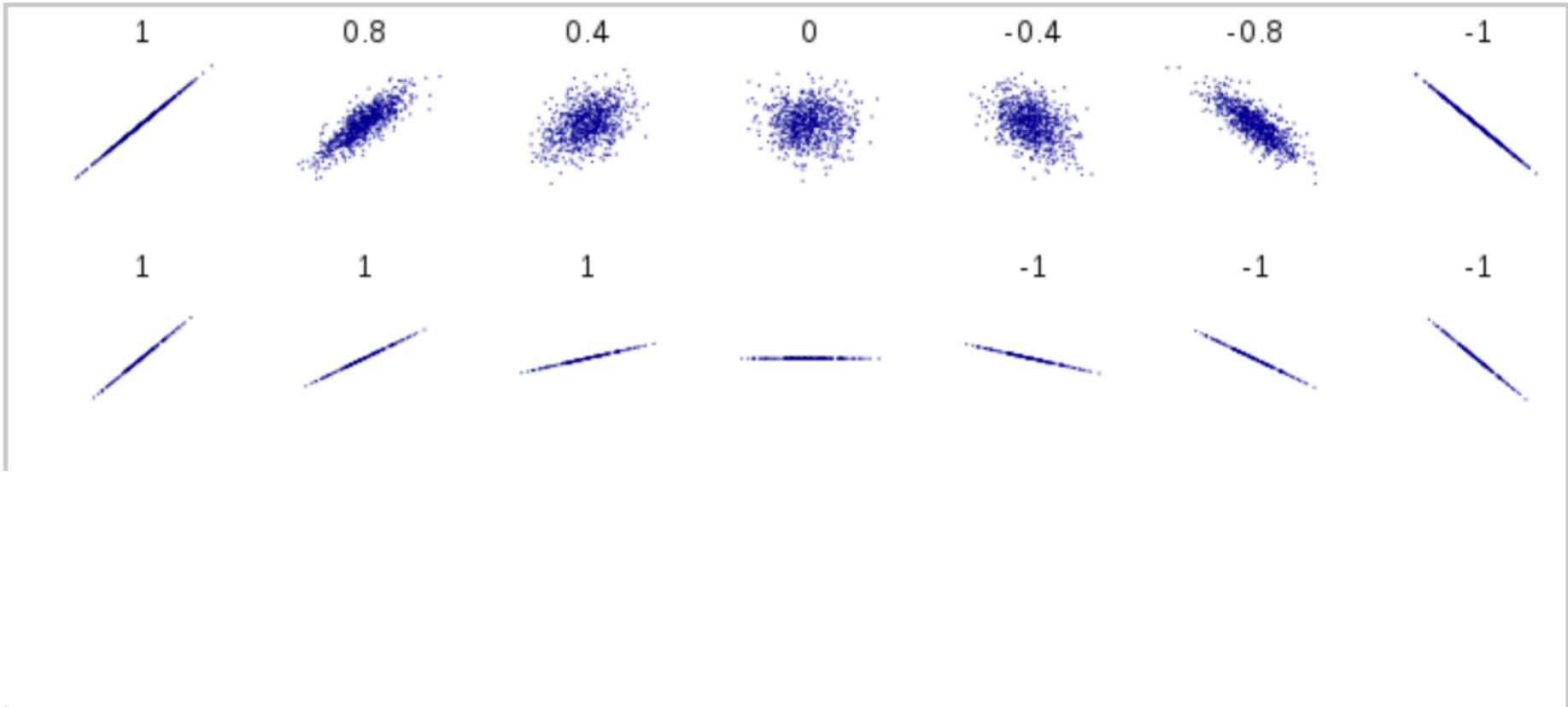
Masters
Écoles d'ingénieurs

DUNOD

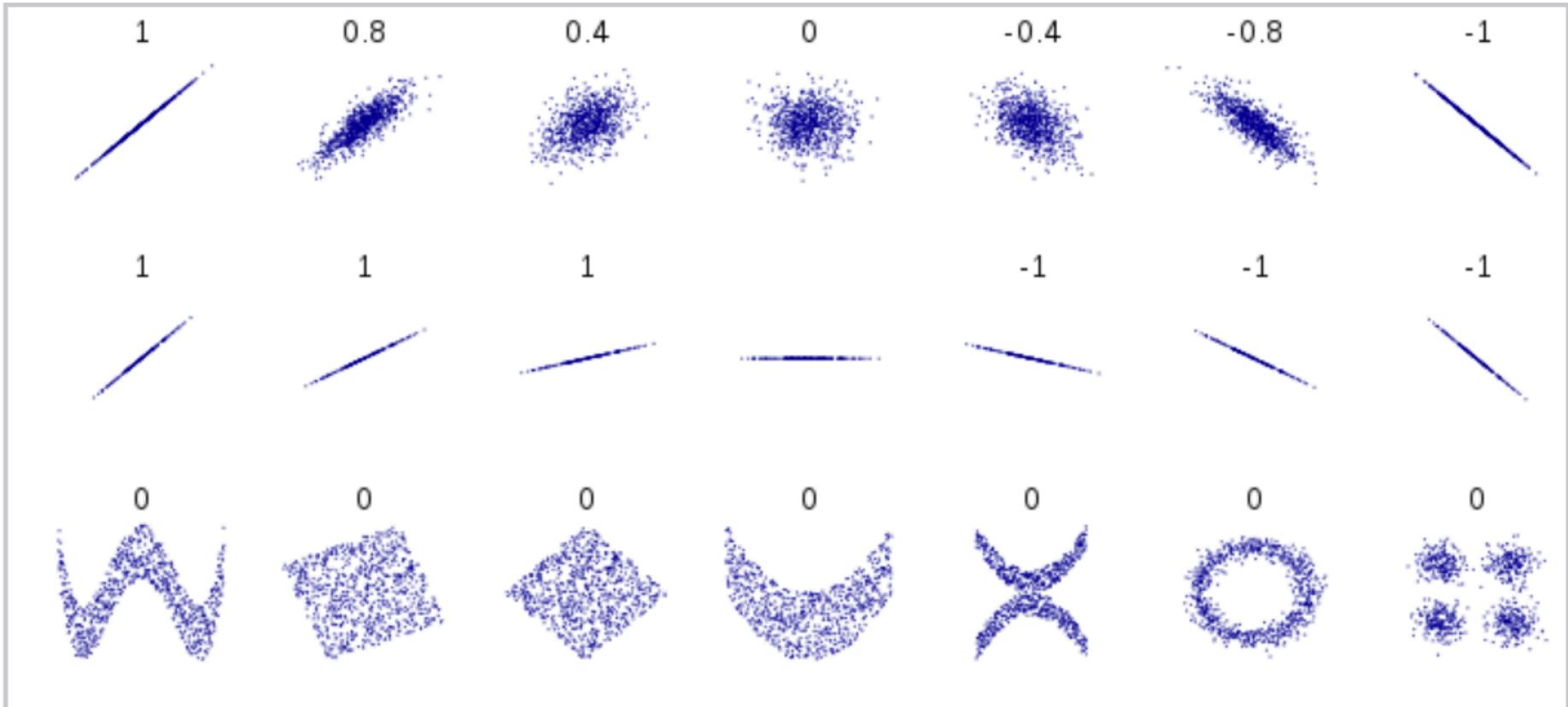




Exemples de coefficients de corrélation : corrélation linéaire pour les deux premières lignes, non linéaire pour la troisième ligne.



Exemples de coefficients de corrélation : corrélation linéaire pour les deux premières lignes, non linéaire pour la troisième ligne.



Exemples de coefficients de corrélation : corrélation linéaire pour les deux premières lignes, non linéaire pour la troisième ligne.



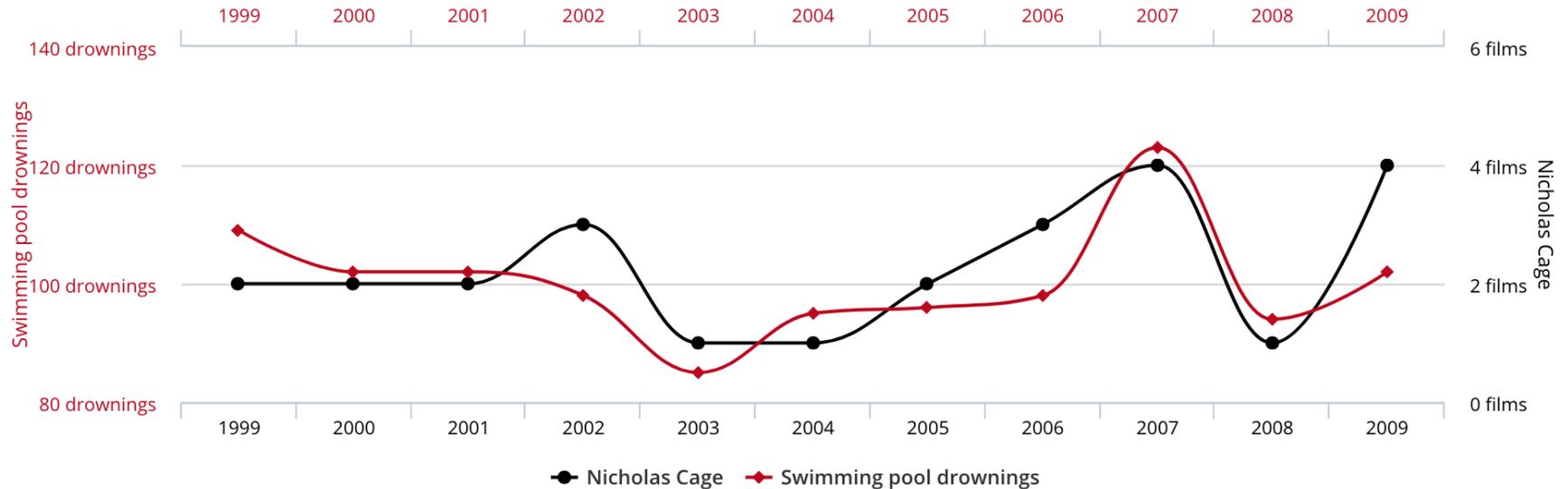
Corrélation et causalité

- Deux variables/mesures sont **corrélées** si
 - Elles évoluent dans le même sens (corrélation positive : quand l'un augmente, l'autre aussi)
 - Ou dans un sens opposé (corrélation négative : quand l'un augmente, l'autre diminue).
- Mais pas forcément de causalité entre ces deux variables/mesures



Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% (r=0.666004)



Data sources: Centers for Disease Control & Prevention and Internet Movie Database

tylervigen.com

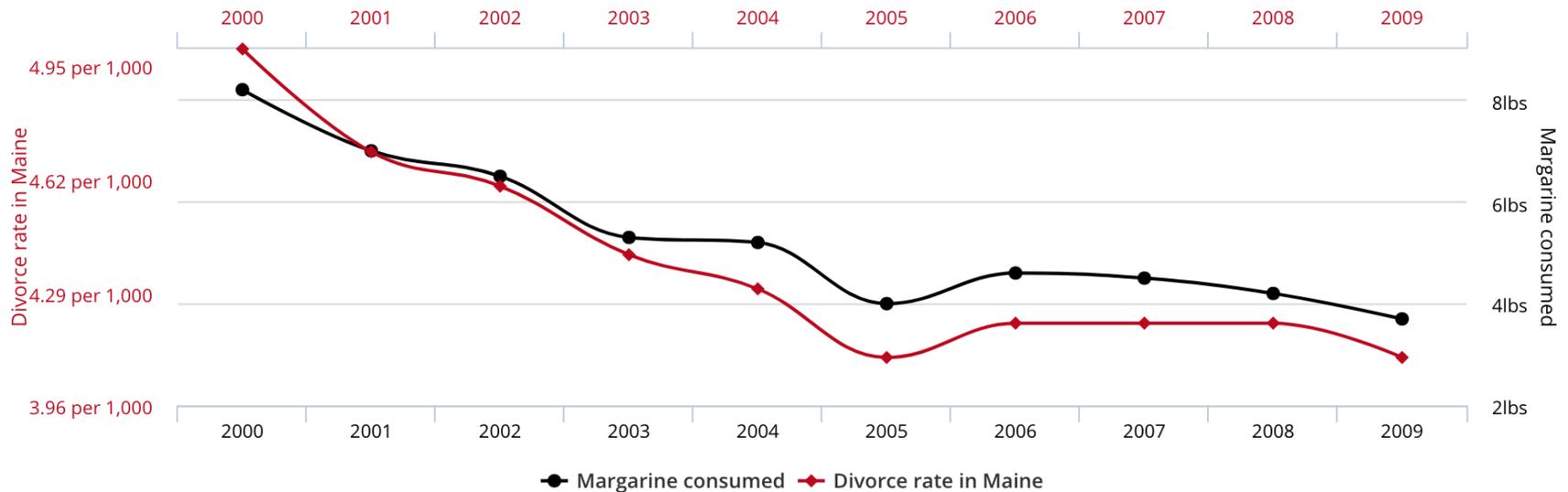


<http://tylervigen.com/spurious-correlations>



Divorce rate in Maine correlates with Per capita consumption of margarine

Correlation: 99.26% (r=0.992558)



tylervigen.com

Data sources: National Vital Statistics Reports and U.S. Department of Agriculture



<http://tylervigen.com/spurious-correlations>



Corrélation et causalité

- Deux variables/mesures sont **corrélées** si
 - Elles évoluent dans le même sens (corrélation positive : quand l'un augmente, l'autre aussi)
 - Ou dans un sens opposé (corrélation négative : quand l'un augmente, l'autre diminue).
- Mais pas forcément de causalité entre ces deux variables/mesures
- La **corrélation** est donc un lien entre deux variables/mesures , sans que pour autant la variation de l'une soit **la cause** de la variation de l'autre.



Variable latente

Espérance de vie et consommation de viande en 2014 par pays

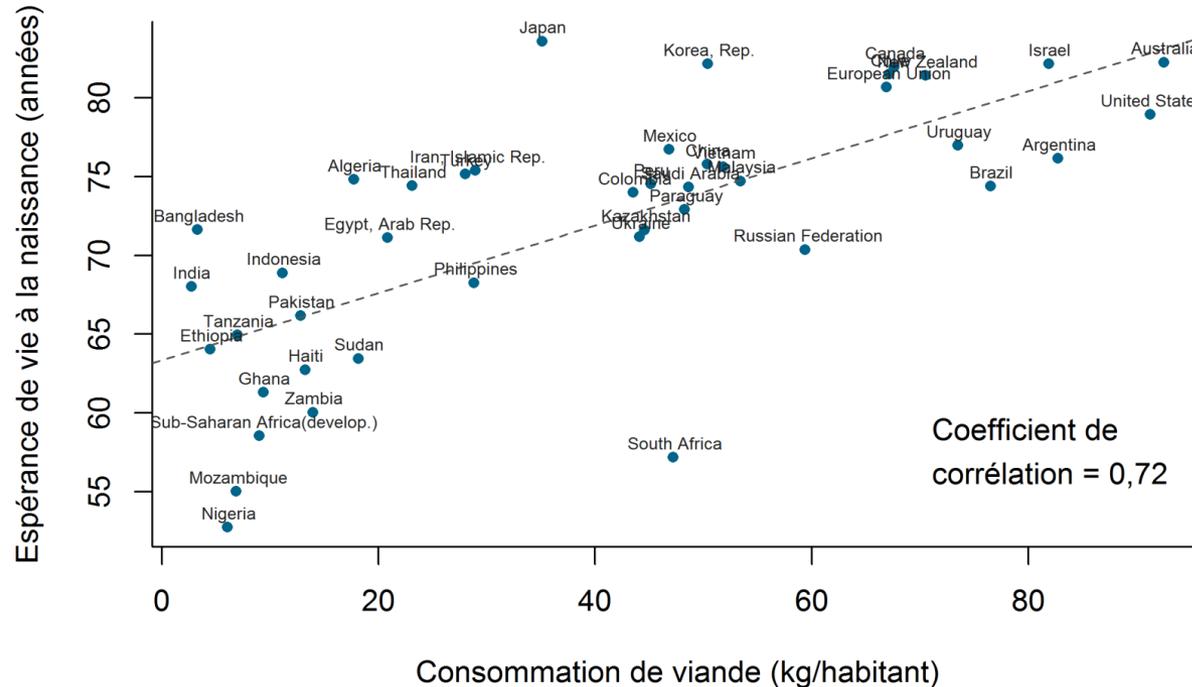


Figure 3 : Espérance de vie à la naissance et consommation de viande en 2014 dans certains pays du monde. Sources : OECD-FAO Agricultural Outlook (Edition 2015) et The World Bank, World Development Indicators.

[https://soepidemio.com/2016/11/14/correlation-statistique-prudence-a-](https://soepidemio.com/2016/11/14/correlation-statistique-prudence-a-interprete/)





Corrélation et dépendances

- Taille et poids ?
 - Quanti*Quanti => R2
- Couleur des yeux et couleurs des cheveux ?
 - Quali*Quali => Chi2
- Couleur des cheveux et taille ?
 - Quali*Quanti =>
 - recoder le quanti en quali (classe de taille)
 - Faire des tests, comparaison de moyennes

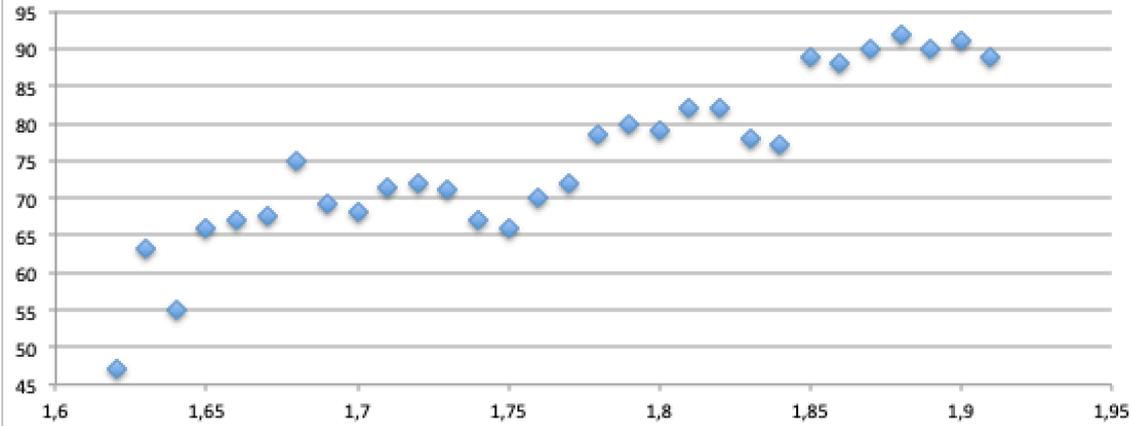


Corrélation et Dépendance

- Corrélation : Variables quantitatives/ordinales
 - Graphes XY
 - R^2 , R de Pearson \Rightarrow variables normales
 - Coefficient de Spearman \Rightarrow non normalité, indicateurs non paramétriques
 - Exemple : Taille et poids
- Dépendance : Variables Qualitatives/nominales
 - Tableau de contingence
 - Calcul du Chi²
 - Exemple : Couleur des yeux et couleur des cheveux



Corrélation Taille Poids



H0 : R=0

H1 : R≠0

R=0,91 - p value <0,05

	blonds	bruns	noirs	roux	Total
bleus	25	9	3	7	44
gris	13	17	10	7	47
marrons	7	13	8	5	33
Total	45	39	21	19	124

Répartition couleur cheveux/yeux

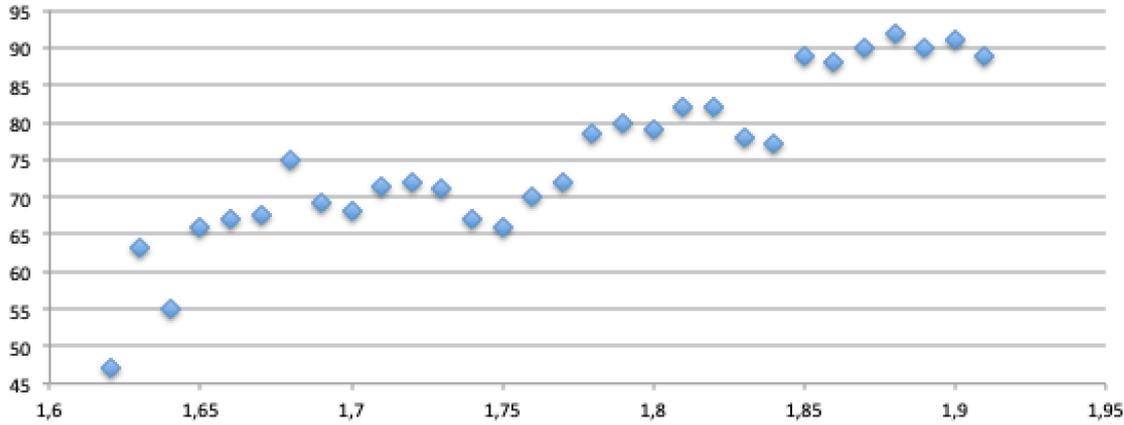
H0 : Variable A est indépendante de la variable B

H1 : Variable A est dépendante de la variable 2

Chi2 =15,1, P value < 0,01



Corrélation Taille Poids



R=0,91 - P value <0,05

H0 : R est égal à 0

H1 : R est significativement différent de 0

	blonds	bruns	noirs	roux	Total
bleus	25	9	3	7	44
gris	13	17	10	7	47
marrons	7	13	8	5	33
Total	45	39	21	19	124

Répartition couleur cheveux/yeux

Chi2 =15,1, P value < 0,01

H0 : la couleur des yeux est indépendante de la couleur des cheveux

H1 : H0 : la couleur des yeux est dépendante de la couleur des cheveux



<https://statistics.ead-minerve.fr>



Matrice de corrélations

<http://www.sthda.com/french/wiki/matrice-de-correlation-avec-r-analyse-et-visualisation>

```
library(Hmisc)
rcorr(as.matrix(mtcars[,1:7]))
```

```
      mpg   cyl  disp    hp  drat    wt   qsec
mpg   1.00 -0.85 -0.85 -0.78  0.68 -0.87  0.42
cyl  -0.85  1.00  0.90  0.83 -0.70  0.78 -0.59
disp -0.85  0.90  1.00  0.79 -0.71  0.89 -0.43
hp   -0.78  0.83  0.79  1.00 -0.45  0.66 -0.71
drat  0.68 -0.70 -0.71 -0.45  1.00 -0.71  0.09
wt   -0.87  0.78  0.89  0.66 -0.71  1.00 -0.17
qsec  0.42 -0.59 -0.43 -0.71  0.09 -0.17  1.00
```

n= 32

P

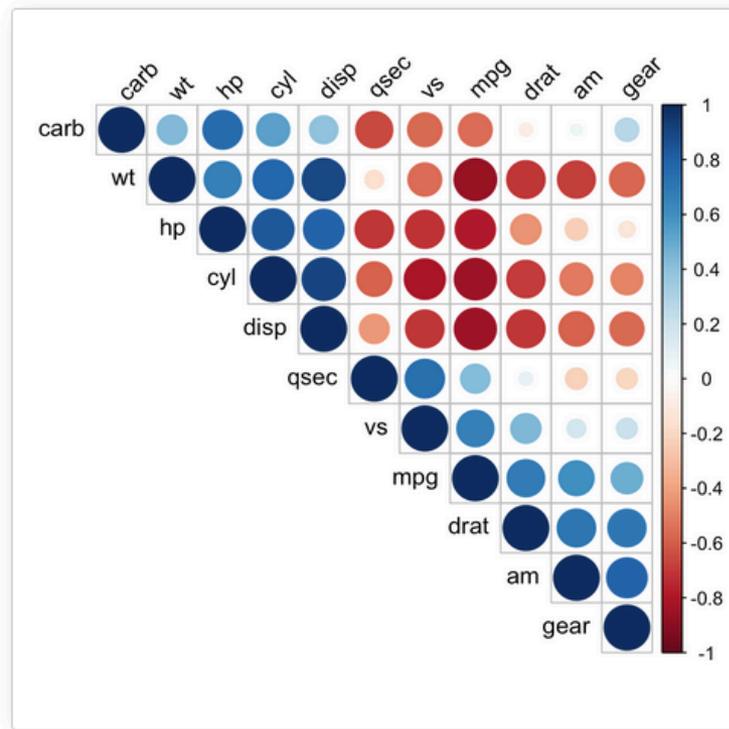
```
      mpg   cyl   disp   hp   drat   wt   qsec
mpg   0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0171
cyl  0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0004
disp 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0131
hp   0.0000 0.0000 0.0000 0.0000 0.0100 0.0000 0.0000
drat 0.0000 0.0000 0.0000 0.0100 0.0000 0.0000 0.6196
wt   0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.3389
qsec 0.0171 0.0004 0.0131 0.0000 0.6196 0.3389 0.0000
```



Matrice de corrélations

<http://www.sthda.com/french/wiki/matrice-de-correlation-avec-r-analyse-et-visualisation>

```
library(corrplot)
corrplot(mcor, type="upper", order="hclust", tl.col="black", tl.srt=45)
```



✓ Les **corrélations positives** sont affichées en bleu et les **corrélations négatives** en rouge. L'intensité de la couleur et la taille des cercles sont proportionnelles aux **coefficients de corrélation**. A droite du **corrélogramme**, la légende de couleurs montre les **coefficients de corrélation** et les couleurs correspondantes.



Tableaux des fréquences

<i>fréquence</i>	blond	bruns	noir	roux	Total
bleu	25	9	3	7	44
gris	13	17	10	7	47
marron	7	13	8	5	33
Total	45	39	21	19	124

<i>%total</i>	blond	bruns	noir	roux	Total
bleu	20,2	7,3	2,4	5,6	35,5
gris	10,5	13,7	8,1	5,6	37,9
marron	5,6	10,5	6,5	4,0	26,6
Total	36,3	31,5	16,9	15,3	100,0

<i>%ligne</i>	blond	bruns	noir	roux	Total
bleu	56,8	20,5	6,8	15,9	100,0
gris	27,7	36,2	21,3	14,9	100,0
marron	21,2	39,4	24,2	15,2	100,0
Total	36,3	31,5	16,9	15,3	100,0