

## Agence nationale de la recherche (ANR) - ANR - Modèle de PGD (français)

### 1. Description des données et collecte ou réutilisation de données existantes

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

#### Recommandations:

- Expliquer quelles méthodologies ou quels logiciels seront utilisés si de nouvelles données sont recueillies ou produites.
- Énoncer les éventuelles restrictions à la réutilisation des données préexistantes.
- Expliquer comment la provenance des données sera documentée.
- Indiquer brièvement le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée mais écartée.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

#### Recommandations:

- Donner des détails sur le type de données : par exemple numérique (bases de données, tableaux), texte (documents), image, audio, vidéo, et/ou médias composites.
- Détailler le format des données : la manière selon laquelle les données sont codées pour le stockage, généralement reflétée par l'extension du nom de fichier (par exemple pdf, xls, doc, txt, ou rdf).
- Justifier l'utilisation de certains formats. Par exemple, les choix d'un format peuvent être guidés par l'expertise du personnel de l'organisme, ou par une préférence pour les formats ouverts, par les standards de format acceptés par les entrepôts de données, par l'usage largement répandu dans une communauté de recherche ou par le logiciel ou l'équipement qui sera utilisé.
- Privilégier les formats standards et ouverts car ils facilitent le partage et la réutilisation à long terme des données (plusieurs catalogues fournissent des listes de ces "formats préférés").
- Donner des détails sur les volumes (qui peuvent être exprimés en espace de stockage requis (octets), et/ou en quantités d'objets, de fichiers, de lignes, et de colonnes).

### 2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

#### Recommandations:

- Indiquer quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données.
- Indiquer quels standards de métadonnées seront utilisés (par exemple DOI, TEI, EML, MARC, CMDI).
- Utiliser les standards de métadonnées des communautés scientifiques lorsque ceux-ci existent.
- Indiquer comment les données seront organisées au cours du projet, en mentionnant par exemple les conventions de nommage, le contrôle de version et les structures des dossiers. Des données bien classées et gérées de façon cohérente seront plus faciles à retrouver, à comprendre et à réutiliser.
- Penser à la documentation qui serait nécessaire pour permettre une réutilisation des données. Il peut s'agir notamment de l'information sur la méthodologie utilisée pour collecter les données, sur les procédures et méthodes d'analyse utilisées, sur la définition des variables, des unités de mesure, etc.
- Tenir compte de la façon dont ces informations seront obtenues et enregistrées par exemple dans une base de données avec des liens vers chacun des fichiers, dans un fichier-texte de type « lisez-moi », dans les en-têtes de fichiers, dans un livre de référence (« code book ») ou dans les cahiers de laboratoire.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

#### Recommandations:

- Expliquer comment la qualité et la conformité de la collecte des données seront contrôlées et documentées. Il s'agit de préciser les processus comme la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la revue par les pairs, ou la représentation basée sur des vocabulaires contrôlés.

### 3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

#### Recommandations:

- Décrire l'endroit où les données seront stockées et sauvegardées au cours du processus de recherche et la fréquence à laquelle la sauvegarde sera effectuée. Il est recommandé de stocker les données dans au moins deux lieux distincts.
- Privilégier l'utilisation de systèmes de stockage robustes, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution d'origine. Le stockage des données sur des ordinateurs portables, des disques durs externes, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

#### Recommandations:

- Expliquer comment les données seront récupérées en cas d'incident.
- Expliquer qui aura accès aux données au cours du processus de recherche et comment l'accès aux données est contrôlé, en particulier dans le cadre de recherches menées en collaboration.

Modèle créé avec DMP OPIDoR. Dernière modification le 14/12/2020

1 of 3

- Tenir compte de la protection des données, en particulier si vos données sont sensibles (par exemple données à caractère personnel, politiquement sensibles des informations ou secrets commerciaux). Décrire les principaux risques et la façon dont ils seront gérés.
- Expliquer quelle politique institutionnelle de protection des données est mise en œuvre.

### 4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel (à sur la sécurité des données sera-t-il assuré ?

#### Recommandations:

Lorsque vous manipulez des données à caractère personnel, veillez à ce que les lois sur la protection des données (par exemple, RGPD) soient appliquées, notamment :

- Obtenir un consentement éclairé pour la préservation et/ou le partage de données personnelles.
- Envisager l'anonymisation des données personnelles pour la préservation et/ou le partage (des données correctement anonymisées ne sont plus considérées comme des données personnelles).
- Envisager la pseudonymisation des données personnelles (la principale différence avec l'anonymisation est que la pseudonymisation est réversible).
- Envisager le chiffrement des données, qui est considéré comme un cas particulier de pseudonymisation (la clé de cryptage doit alors être stockée séparément des données, par exemple par un tiers de confiance).
- Expliquer si une procédure d'accès spécifique a été mise en place pour les utilisateurs autorisés à accéder aux données personnelles.

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

#### Recommandations:

- Expliquer qui sera le propriétaire des données, c'est-à-dire qui aura le droit d'en contrôler l'accès :
- Expliquer quelles conditions d'accès s'appliqueront aux données. Les données seront-elles librement accessibles, ou des restrictions seront-elles appliquées ? Si oui, lesquelles ? Envisager l'utilisation de licences concernant l'accès et la réutilisation des données.
- S'assurer de couvrir, dans l'accord de consortium, ces questions de droits de contrôle d'accès aux données pour les projets multipartenaires et en cas de propriété partagée des données.
- Indiquer si les droits de propriété intellectuelle (par exemple la directive bases de données, droits *sui generis*) sont affectés. Dans l'affirmative, expliquer lesquels et comment cela sera traité.
- Indiquer s'il y a des restrictions sur la réutilisation des données fournies par des tiers.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

#### Recommandations:

- Déterminer si les questions d'éthique auront une incidence sur la façon dont les données seront stockées et transférées, qui pourra les voir ou les utiliser et quelles durées de conservation leur seront appliquées. Démontrer que ces aspects sont bien pris en compte et planifiés.
- Adopter les codes de conduite nationaux et internationaux et le code d'éthique institutionnel et vérifier si une revue des pratiques (par exemple par un comité d'éthique) est requise pour ce qui concerne la collecte de données dans le cadre du projet de recherche.

### 5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

#### Recommandations:

- Expliquer comment les données pourront être retrouvées et partagées (par exemple, par le dépôt dans un entrepôt de données de confiance, l'indexation dans un catalogue, par l'utilisation d'un service de données sécurisé, par le traitement direct des demandes de données, ou l'utilisation de tout autre mécanisme).
- Définir le plan de préservation des données et fournir l'information sur la durée d'archivage prévue des données.
- Expliquer à quel moment les données seront rendues disponibles. Indiquer les délais de publication prévus. Expliquer si une utilisation exclusive des données est revendiquée et, dans l'affirmative, pour quelle raison et pour quels motifs. Indiquer si le partage des données sera différé ou limité, par exemple pour des raisons de publication, pour protéger la propriété intellectuelle ou le dépôt de brevets.
- Indiquer comment vous envisagez d'utiliser les données. S'il s'avère nécessaire de restreindre l'accès pour certaines communautés ou d'imposer un embargo sur le partage de données, expliquer comment et pourquoi. Expliquer les mesures qui seront prises pour dépasser ou minimiser ces restrictions.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

#### Recommandations:

- Indiquer quelles données ne doivent pas être divulguées ou doivent être détruites pour des raisons contractuelles, légales, ou réglementaires.
- Indiquer comment il sera décidé quelles données garder. Décrire les données qui seront à préserver à long terme.
- Décrire les utilisations (et/ou les utilisateurs) prévues des données dans un cadre de recherche.

Modèle créé avec DMP OPIDoR. Dernière modification le 14/12/2020

2 of 3

- Indiquer où les données seront déposées. Si aucun entrepôt reconnu n'est proposé, démontrer dans le plan de gestion des données que les données pourront être prises en charge efficacement au-delà de la durée de financement du projet. Il est recommandé de démontrer que les politiques des entrepôts et les procédures de dépôts (y compris les standards de métadonnées, et coûts mis en œuvre) ont été vérifiées.

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

*Recommandations:*

- Indiquer si les utilisateurs potentiels auront besoin d'outils spécifiques pour l'accès et la (ré)utilisation des données. Tenir compte de la durée de vie des logiciels nécessaires pour accéder aux données.
- Indiquer si les données seront partagées via un entrepôt, si les demandes d'accès seront traitées en direct, ou si un autre mécanisme sera utilisé.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

*Recommandations:*

- Expliquer comment les données pourraient être réutilisées dans d'autres contextes. Les identifiants pérennes devraient être appliqués de manière à ce que les données puissent être localisées et référencées de façon fiable et efficace. Les identifiants pérennes aident aussi à comptabiliser les citations et les réutilisations.
- Indiquer s'il sera envisagé d'attribuer aux données un identifiant pérenne. Typiquement, un entrepôt pérenne de confiance attribuera des identifiants pérennes.

## 6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

*Recommandations:*

- Décrire les rôles et les responsabilités concernant les activités de gestion des données, par exemple : saisie des données, production des métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données. Nommer la(les) personne(s) responsable(s) impliquée(s) dans la mesure du possible.
- Pour les projets menés en collaboration, expliquer comment s'effectue la coordination des responsabilités de gestion des données entre partenaires.
- Indiquer qui est responsable de la mise en œuvre du PGD, et qui s'assure qu'il est examiné et, si nécessaire, révisé.
- Envisager des mises à jour régulières du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

*Recommandations:*

- Expliquer comment les ressources nécessaires (par exemple le temps) à la préparation des données pour le partage/préservation (curation des données) ont été chiffrées. Examiner et justifier soigneusement toutes les ressources nécessaires pour diffuser les données.
- Il peut s'agir de frais de stockage, de coût matériel, de temps de personnel, de coûts de préparation des données pour le dépôt, de frais d'entrepôt et d'archivage.
- Indiquer si des ressources supplémentaires sont nécessaires pour préparer les données en vue de leur dépôt ou pour payer tous les frais demandés par les entrepôts de données. Si oui, précisez le montant et comment ces coûts seront couverts.



## « 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.

### Nom de l'expérimentation

- Donner un nom à l'expérimentation (cela permet d'avoir un moyen mnémorique pour se rappeler de quelle expérimentation il s'agit, du lieu des circonstances, des outils testés, des participants, etc.)

### Objectifs de l'expérimentation

- Quels sont les objectifs de l'expérimentation ?
- Quelles sont les contributions scientifiques concernées par l'expérimentation et quel est leur état d'avancement (à explorer, à élaborer, à construire, à évaluer, etc.) ?
- En quoi, l'expérimentation fera évoluer la contribution scientifique ?
- Quels sont les outils associés qui sont concernés par cette expérimentation (e.g. maquette, site web, robot, etc.) ?
- S'il y en a, quels sont les composants (parties des outils) concernés par l'expérimentation ? (l'objectif est de décomposer l'outil en composants pour identifier les points précis à construire ou évaluer)
- Quel est l'état d'avancement de l'outil et des composants : inexistant (e.g. l'outil n'existe pas encore le participant va contribuer à sa conception), numérique (e.g. plateforme numérique, robot, etc.) ou statique (e.g. format papier, livre, etc.) ?

### Questions ou hypothèses et mesures

- Quelles sont les questions et /ou les hypothèses qui devront trouver des éléments de réponses lors de cette expérimentation ?
- Quelles sont les données existantes que vous pouvez mobiliser ? (e.g. des corpus de données, des traces d'activités, des documents, des archives, etc.)
- Quelles sont les mesures qui seront recueillies lors de l'expérimentation ? (e.g. vitesse, performance, activités, attentes, besoins, changements de pratiques, satisfaction, etc.)
- Pouvez-vous vous appuyer sur des questionnaires ou des guides d'entretien publiés dans la littérature ? Lesquels ? (e.g. questionnaire pour l'utilisabilité SUS de Brooke)

## Si l'expérimentation implique des humains, description des participants (information à reprendre du guide pour l'arbre de décision)

- Quel est le profil des participants à l'expérimentation (e.g. profession, niveau d'expertise, âge, etc.) ?
- Pourquoi avez-vous choisi ces profils ?
- Quels sont les liens entre les participants ? se connaissent-ils ?
- En fonction des mesures et des données que vous souhaitez récolter, que connaissez-vous de vos participants ? (e.g. - je ne connais pas leur manière d'enseigner la chimie, - je connais très bien leur manière de gérer les projets, - je ne connais pas les habitudes des familles dans l'utilisation de la domotique, - j'ai déjà fait des observations in situ, je connais les pratiques de mes utilisateurs avec les outils connectés)
- Pourquoi ces participants sont-ils mobilisés ? (e.g. observer en situation réelle, discuter avec eux, quantifier leurs pratiques, leur faire confronter des idées entre eux, etc.)
- Comment le participant pourra-t-il utiliser des outils ou des composants pendant les expérimentations ?
- Quelles seront les tâches demandées aux participants ?
- Comment présenterez-vous ces tâches aux participants ?
  - CONSEIL 1 : A priori, les participants ne sont pas experts de votre domaine, il n'est pas nécessaire de leur présenter votre contribution scientifique.
  - CONSEIL 2 : Pour répondre à cette question, il faut vous mettre à la place du participant. A priori, il n'a aucune connaissance sur ce qu'il doit tester, il faut lui présenter clairement l'outil ou le composant sans pour autant lui révéler vos hypothèses ou questions.
- Quelles consignes donnerez-vous à vos participants ?
- Combien de participants sont pressentis ? et combien après le recrutement ?
- Où se passe l'expérimentation ? (e.g. in lab, in situ, zoom, etc.)
- Les participants sont-ils consultés seuls ou en groupe ?
- Comment procédez-vous au recrutement des participants ?
  - CONSEIL 1 : *Le recrutement peut être long, penser à l'anticiper*
  - CONSEIL 2 : *Rédiger un document de consentement*

## RGPD et comité d'éthique

- Quelles sont les démarches à réaliser auprès de la RGPD pour le traitement des données ?
- Quelles sont démarches à réaliser auprès d'un comité d'éthique pour déclarer l'expérimentation ?



## Méthodes et outils de production des données

Les outils de production de données permettent de collecter les données lors d'une expérimentation.

### Méthodes de production de données

- A l'aide du logigramme (MATUI), indiquer les méthodes de production de données que vous utiliserez.

### Outils de production de données existants

- Quels sont les outils que vous utiliserez pour produire des données (e.g. un capteur de traces, un oculomètre, etc.) ?
- Quels sont les outils de production de données qui existent dans la littérature (e.g. des questionnaires, des grilles d'observation, etc.) ? Lesquels utiliserez-vous ? (indiquer les références)

### Outils de production de données à produire

- Quels sont les outils de production de données que vous créerez ? (e.g. questionnaire, guide d'entretien, etc.) ?
- Quel est le matériel expérimental à construire pour réaliser l'expérimentation ? (e.g. jeux de données pour un benchmark, présentation, des schémas etc.)

### Données produites

- Quel est le matériel et quelles sont les données qui seront produits lors de la passation de l'expérimentation ? (e.g. fichiers, schéma, audio, traces, etc.)
- Quel est le format des traces à produire si un outil numérique est utilisé ? (e.g. fichier csv, norme XAPI, etc.)

### Matériel technique

- Quel est le matériel technique nécessaire à avoir pour la capture des données (e.g. densitomètre, capteurs, balance, caméra, enregistreur, agitateur magnétique.etc.) ?
- Conseil : prévoir un matériel de qualité pour les enregistrements et des batteries chargées.

### Outils d'analyse de données

- Quelles sont les méthodes d'analyse pressenties pour analyser les données ?
- Quels sont les outils pressentis pour analyser les données ?
- Quels sont les traitements à réaliser pour valider la qualité des données ?
- Quel plan de traitement des données suivrez-vous ? (se référer au data management plan (DMP) de votre institution)

### Planning

- Quelles sont à gros grain les étapes de l'expérimentation ?



## « Organiser une expérimentation »

**Quand** : Ce document doit être fait lors de la construction de son expérimentation et testé avec une expérience pilote.

**Pourquoi** : Il permet d'organiser le déroulement, de documenter et de tracer une expérimentation qui comprend plusieurs étapes. Il permet de la structurer.

**Apports pour la rédaction de thèse** : Il servira à la rédaction de la partie méthode de la thèse. Il sera joint aux annexes de la thèse.

Ce guide peut servir lorsqu'une expérimentation comporte plusieurs étapes (*e.g. un test utilisateur avec un entretien qualitatif avant et après l'utilisation d'un dispositif*). Ce guide liste les éléments qui doivent être décrits avec précision et il fournit un exemple de tableau pour organiser une expérimentation ou plusieurs.

### Le guide est structuré en étape

- **No d'étape** : Indiquer un numéro d'étape
- **Objectif de l'étape** : Décrire les objectifs de l'étape. Pour les étapes qui concernent les composants, utiliser les trois verbes d'action de la démarche centrée utilisateur (explorer, co-construire, évaluer) (Mandran et al., 2013)
- **Hypothèses et/ou questions testées dans cette étape** : Indiquer les questions ou les hypothèses expérimentales qui devront trouver des éléments de réponses lors de l'étape de cette expérimentation. Ces questions et hypothèses correspondent à celles posées dans le protocole expérimental.
- **Activités réalisées** : Lister les activités faites pendant cette étape. (*e.g. présentation du modèle, tests utilisateur, lecture du dictionnaire*)
- **Participants concernés** : Utilisateurs, Enseignants, Chirurgiens, Opérateurs, ...
- **Outils ou composant(s) associé(s)** : Indiquer le(s) outils et/ou composant(s) associé(s) qui seront testés dans l'étape.
- **Etat de l'outil ou du composant** : Indiquer l'état d'avancement de(s) composants et dans quelle forme il va être utilisé par les participants (numérique/statique)
- **Personnes en charge du (des) composant(s)** : Inscrire les noms des personnes en charge de concevoir, développer ou évaluer le(s) composant(s)
- **Matériel expérimental** : Lister les matériels expérimentaux nécessaires à cette étape (*e.g. guide d'entretien, d'annotation*). Indiquer le rôle de ces documents (présentation, recueil).
- **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** : Cumul de la durée des activités

Remarque :

Vous pouvez modifier le guide comme vous le souhaitez en ajoutant par exemple une colonne remarque pour annoter les événements ou réflexions qui surviennent sur les différentes étapes de l'expérimentation. Pour documenter et tracer le déroulement de cette expérimentation, le guide peut rester ouvert pendant l'expérimentation et à la suite de celle-ci.

Référence :

Mandran, N., Dupuy-Chessa, S., Front, A., & Rieu, D. (2013). Démarche centrée utilisateur pour une ingénierie des langages de modélisation de qualité. *Ingénierie des systèmes d'information*, 18(3), 65-93. <https://doi.org/10.3166/isi.18.3.65-93>

## Tableau d'organisation

(pour alléger la présentation les durées ne sont pas indiquées sur cet exemple)

Nom de la personne en charge de l'expérimentation :

Date et heure de l'expérimentation :

Lieu de la passation :

No d'étape	Objectif de l'étape	Hypothèses et questions en lien avec cette étape	Activités réalisées	Participants concernés	Outils ou composant(s) associé(s)	Etat de l'outil ou du composant	Personnes en charge du (des) composant(s)	Matériel expérimental
1	Présentation de l'expérimentation et du guide de conduite de la recherche	/	Présentation par le chercheur	Doctorants/ Directeurs de thèse	/	/	/	Powerpoint (présentation)
2	Atelier : Personnalisation du guide de conduite de la recherche	Les trois spécificités du guide (tâches, indicateurs, rôles/acteurs) sont-elles utilisables par rapport à un travail de thèse ? (Q3)	Test utilisateur	Doctorants/ Directeurs de thèse	Guide "méthode cyclique THEDRE, et liste des tâches"	Statique/ format papier	Estelle	Feuille A3 / enveloppe avec les 60 tâches / gommettes / Post-It
3	Atelier : Echange autour de l'utilisation du guide de conduite de la recherche	Les trois spécificités du guide (tâches, indicateurs, rôles/acteurs) sont-elles utilisables par rapport à un travail de thèse ? (Q3)	Focus group	Doctorants/ Directeurs de thèse	Guide "méthode cyclique THEDRE, et liste des tâches"	Statique/ format papier	Estelle	Enregistreur (dictaphone + caméra)
...								



## « 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.

### Informations générales sur l'expérimentation

- Quel est le nom de l'expérimentation ? (fait référence au nom utilisé dans le protocole expérimental)
- Quels sont la date et le lieu de passation ?
- Quel est le nom du/des chercheur(s) qui ont fait la passation de l'expérimentation ?
- Quelle note attribuez-vous à la passation ? (selon le niveau de pertinence ou d'intérêt par rapport à la question de recherche/hypothèse posée. Les notes vont de 0 à 5 : 0 non pertinent à 5 très pertinent) Lors de l'analyse, cette notation permet au chercheur d'identifier les données les plus pertinentes pour la rédaction des résultats.

### Description de l'expérimentation

- Quel nom souhaiteriez-vous donner à cette passation pour illustrer ce qui s'est passé ? Donner un nom différent de celui de l'expérimentation et qui reflète ce qui s'est passé pendant la passation, est un bon moyen de se remémorer la situation lors de la rédaction des résultats.
- Comment les interactions avec les participants ou entre les participants se sont passées ? Cela permettra d'indiquer l'ambiance générale que vous avez ressentie pendant la passation.
- Quels événements pouvant biaiser les résultats, se sont produits pendant la passation ? (*e.g. le déclenchement d'une alarme incendie qui stoppe le focus-group*)
- Quelles réponses apporteriez-vous à vos questions ou hypothèses suite à la passation ? Résumer ce qui a été dit par le(s) participant(s) ou ce que vous avez pu relever pendant la passation et qui répond à ces questions ou hypothèses.
- Quels sont les éléments novateurs qui sont abordés dans cette passation et qui contribuent à l'amélioration de la connaissance scientifique ou des outils associés ?

### Bilan de l'expérimentation

- Quelles étapes de l'expérimentation se sont bien déroulées et quel matériel expérimental était bien dimensionné pour la passation ? (i.e. les points forts).
- Quelles étapes de l'expérimentation ne se sont pas bien déroulées et pourquoi ? Quel matériel expérimental a posé problème et pourquoi ? (i.e. les points faibles).
- Quels sont les points d'améliorations à apporter à l'expérimentation ?



## « 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 ?

### Méthode de la littérature

- Quelles sont les méthodes d'analyse utilisées dans la littérature ? (*e.g. visualisation, ANOVA, modélisation, analyse thématique, etc.*)
- Quels sont les outils, logiciels utilisés dans la littérature ? (*e.g. R, SPSS, python, CAQDAS, etc.*)
- Pourquoi allez-vous utiliser ces méthodes d'analyse ?
- Sinon, pourquoi n'utilisez-vous pas ces méthodes d'analyse ?
- Finalement, quelles méthodes, quels logiciels allez-vous utiliser ?

### Pré-traitement des données, quelques conseils

Le **pré-traitement des données** permet de vérifier la qualité des données avant de faire les traitements. Pour cela, il est recommandé d'étudier la distribution des différentes variables :

- Calcul des fréquences, moyennes, médianes, écart types, min, max, quartiles.
- Faire les histogrammes et des box plots pour les variables numériques afin de détecter les données aberrantes.
- Faire des diagrammes en bâton pour les variables nominales, ordinales afin de détecter les données aberrantes.

Pour le travail sur la **qualité des données**, ouvrage à utiliser : Di Ruocco, & al. "la qualité des données concepts de base et techniques d'amélioration" in L.Berti Equille (dri.) la qualité et la gouvernance des données, Hermès-Lavoisier, Paris 2012

- Décrire le prétraitement que vous allez faire.
- Rédigez ce que vous observez sur ces pré-traitements. (Il y a toujours beaucoup d'informations dans ces pré-traitements)
- Si vous avez observé des données aberrantes, qu'avez-vous fait de ces données ? Pourquoi ?



- Ces pré-traitement vous ont-ils permis d'enrichir vos questions de recherche ?
- Ces pré-traitement vous ont-ils permis de créer de nouvelles variables pour répondre à vos questions de recherche (*e.g. combinaison de variables, recodage, etc.*) ? Lesquelles ?

### **Traitement des données**

- Décrire le traitement que vous allez faire ? (*e.g. méthodes, logiciels, langage, etc.*)
- Quelles sont les références bibliographiques sur ces méthodes et logiciels ?
- Quelles sont les représentations graphiques qui vont faire parler vos données ?

### **Analyses et résultats**

- Identifier les traitements de données que vous allez conserver pour répondre à votre question de recherche ?
- Rédigez ce que vous observez sur ces traitements.
- En quoi ces traitements répondent ou non à vos questions de recherche ?

### **Limite du traitement des données**

- Indiquez les limites de ce traitement de données.
- Indiquez comment il aurait pu être amélioré.

Procédures et Indicateurs pour la qualité des données – Janvier 2017

Les données produites lors d'une expérimentation fournissent à la fois des fichiers de données qui sont analysés pour produire des résultats et des métadonnées qui décrivent le contenu des fichiers de données. Les approches pour contrôler la qualité des données et les indicateurs vont se préoccuper des métadonnées et des données. [Berti-Equille 2007] propose 4 approches pour contrôler la qualité des données : (1) préventive, (2) adaptative, (3) corrective et (4) diagnostique. L'approche préventive permet un contrôle en amont avant la production des données (p.ex., un test de production des données par un capteur garantira que les données produites en temps réel sont correctes). L'approche adaptative permet la vérification des données en temps réel (p.ex.: pendant la capture de données une application permet d'identifier les données aberrantes, par exemple une augmentation soudaine de la température sur un capteur). **Les approches diagnostiques et correctives sont menées après la production de données. L'approche diagnostique comprend la comparaison avec la réalité sur le terrain et la gestion des métadonnées. L'approche corrective comprend, entre autres, la correction par rapport à la réalité du terrain, l'imputation de données manquantes, le redressement et l'élimination des doublons.**

Afin d'être opérationnelles, les 4 approches pour suivre la qualité des données doivent s'accompagner d'indicateurs de qualité. [Di Ruocco et al. 2012] définissent dix indicateurs :

1. **Pertinence** : capacité des données répondre aux besoins actuels et futurs des utilisateurs.
2. **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).
3. **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).
4. **Accessibilité** : la facilité de localisation et d'accès aux données et aux métadonnées.

5. **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.
  
6. **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.
  
7. **Cohérence** : absence d'informations conflictuelles. (p.ex., l'âge des enfants doit être inférieurs à celui de leurs parents).
  
8. **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).
  
9. **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 »).
  
10. **Consistance** : Quand une entité est recopiée, il y a consistance si on retrouve les mêmes valeurs d'attributs dans toutes les bases.