



# Collecte des données de recherche

Identifier les jeux de données générés  
ou utilisés durant le projet de  
recherche

# Prérequis



Identifier le **type de données** utilisé:

Données produites  
Données recueillies  
Données **réutilisées**



Identifier le degré de sensibilité des données



Identifier le(s) format(s)

Au moment de la collecte  
Au cours du projet de recherche  
Pour la conservation des fichiers après l'achèvement du projet de recherche

# Transposition au projet Héritage Canada

- Données préexistantes, recueillies et **réutilisées** dans le cadre de la recherche
- Données libres d'utilisation grâce à **l'octroi** d'une licence
- Format des fichiers:
  - Initialement en XML
  - Durant le projet de recherche
    1. Importées pour lecture en format .xlsx
    2. Nettoyage et traitement des données grâce à des logiciels appropriés
    3. Exportées après traitement en format .csv
    4. Format final: projection cartographique



# Structure et gestion des fichiers



Déterminer une méthode de  
nommage des fichiers



Conserver une trace des  
différentes copies et versions  
des fichiers



Assurer le suivi du  
changements des formats et  
logiciels de traitement des  
données

# Transposition au projet Héritage Canada

- Exemple de **procédure de nommage** des fichiers:
  1. Une fois le fichier .xlsx obtenu, nous utiliserons le logiciel **OpenRefine** pour « nettoyer » les données et ne conserver que les données des attributs **@placeName**, **@long** et **@lat** de la **balise <place>**, (...)
  2. Nous utiliserons ensuite les attributs **@offset** et **@length** de la **balise <TextLine>**, qui permettent de récupérer l'objet (la chaîne de caractères précise dans la ligne) balisée.
- Les noms des lieux balisés sont donc conservés avec leur orthographe original issu des transcriptions.
- L'identité de chacun des documents auxquelles ces entités sont localisées ainsi que la longitude et la latitude qui leur sont associées.





Questions ?