# Synthèse des étapes 2 et 3

Date de génération : 30/05/2025 20:35:02

## Étape 2 – Extraction des QID

Cette étape utilise une requête SPARQL sur le label Wikidata pour récupérer le QID associé à chaque titre extrait à l’étape 1.  
- Requête SPARQL : SELECT DISTINCT ?item WHERE { ?item rdfs:label "Titre"@fr. }  
- Résultat : QID (ex: Q123456) est injecté dans l’objet EntreeHistorique.  
- Données enrichies : QID uniquement.  
- Fichier de sortie : step2\_JD01\_batch\_001.json  
- Répertoire : step2\_coords/  
- Format : JSONL structuré avec EntreeHistorique (titre, url, qid, process)

## Étape 3 – Enrichissement SPARQL (P625, P31, nb\_langues)

Cette étape interroge Wikidata par QID pour récupérer :  
- les coordonnées GPS (P625)  
- le type principal (P31)  
- le nombre de langues ayant un article (via schema:isPartOf)  
  
Les coordonnées sont projetées en Lambert 93.  
Un filtre est appliqué pour conserver uniquement :  
- Latitude entre 40° et 51°  
- Longitude entre -6° et 11°  
  
Les données rejetées (hors zone ou sans P625) sont loggées.  
- Données enrichies : lat, lon, x\_l93, y\_l93, p31, nb\_langues, note (auto)  
- Fichier de sortie : step3\_JD01\_batch\_001.json  
- Répertoire : step3\_summary/  
- Format : JSONL structuré avec EntreeHistorique enrichi