

## Petit (2.) point sur le traitement des annuaires

Julie Gravier

9 mai 2022

- Actuellement : sélection de 8 dates clefs car au départ les listes du géocodage n'avaient pas le même ordonnancement que le ner xml (dans l'idée de découper à postériori les entrées multiples), à savoir 1839 (Lamy), 1845, 1855 (Didot), 1864, 1875, 1885, 1893 et 1904 (Didot Bottin) ;
- Test découpe des entrées à postériori sur le Didot-Bottin 1870a et lier le géocodage aux nouvelles entrées créées. Faisable mais pas très pratique : où effectuer la découpe dans ner xml ? Le géocodage est un ensemble d'objets et non une array.
- Circuit des données vers les utilisateurs : quels formats ? Ajouter "title" et "activity" comme une array en dehors de ner xml, comme pour les "persons" dans les JSON ?

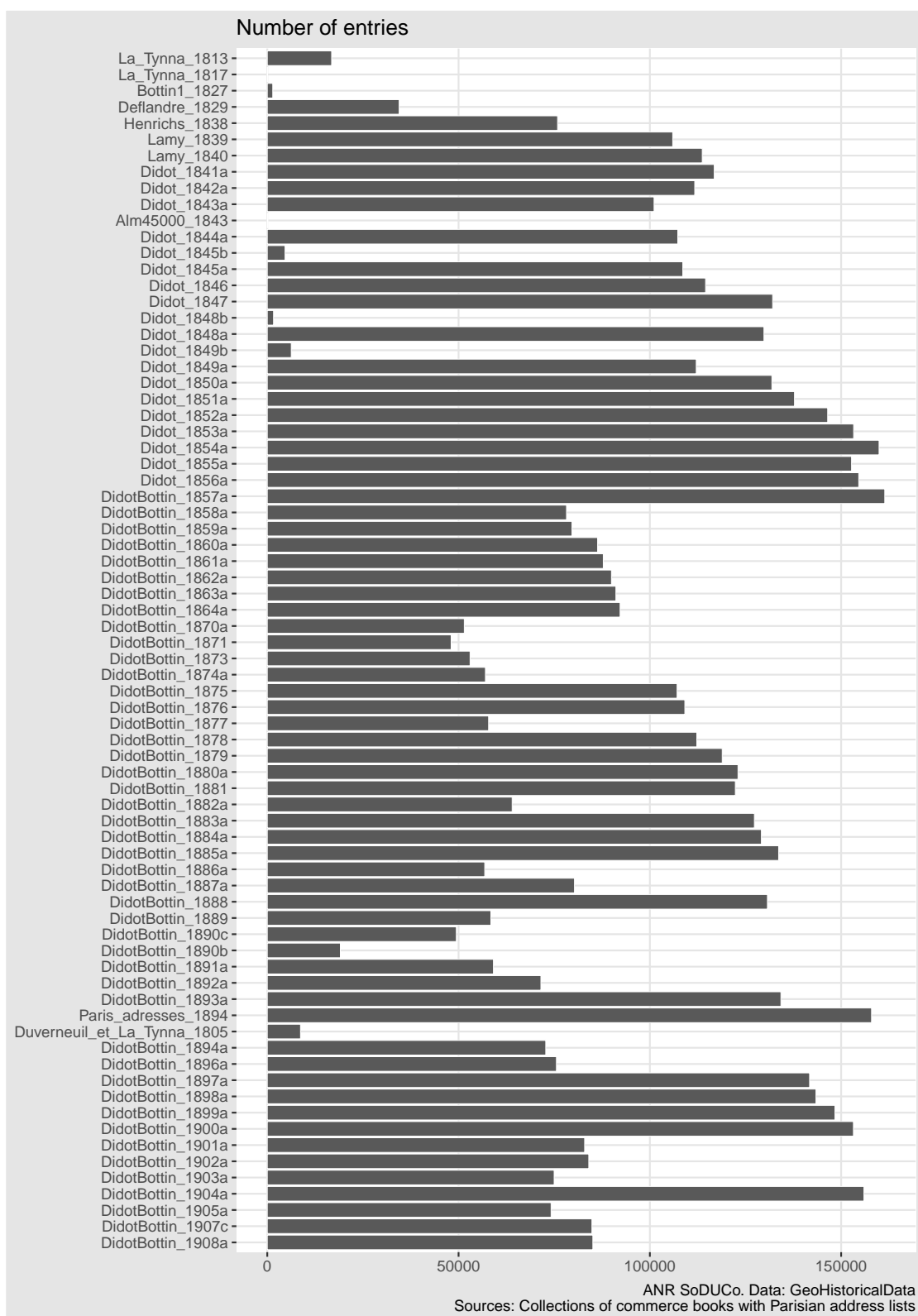


FIGURE 1 – Les entrées par annuaire

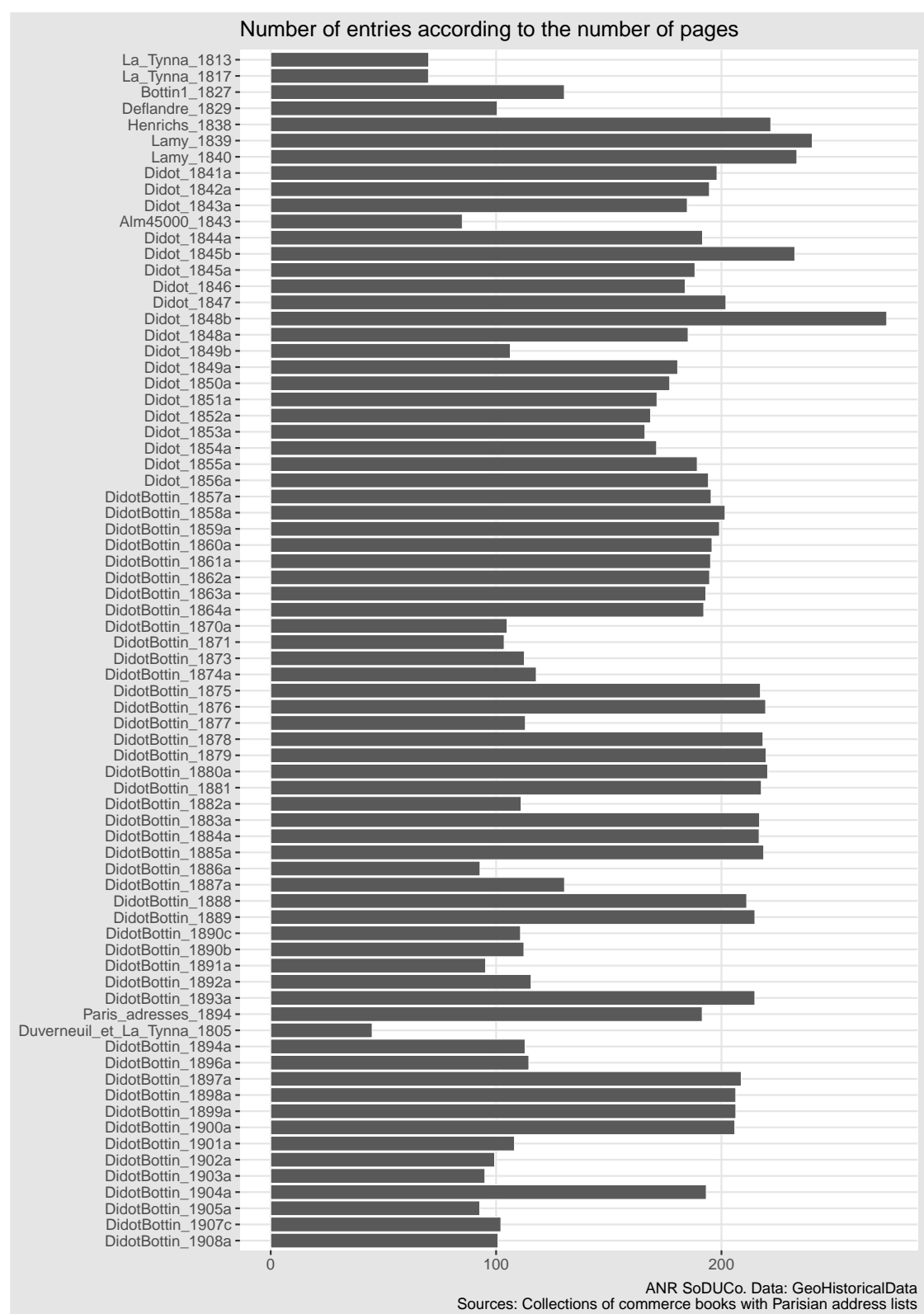


FIGURE 2 – Les entrées rapportées au nombre de pages par annuaire

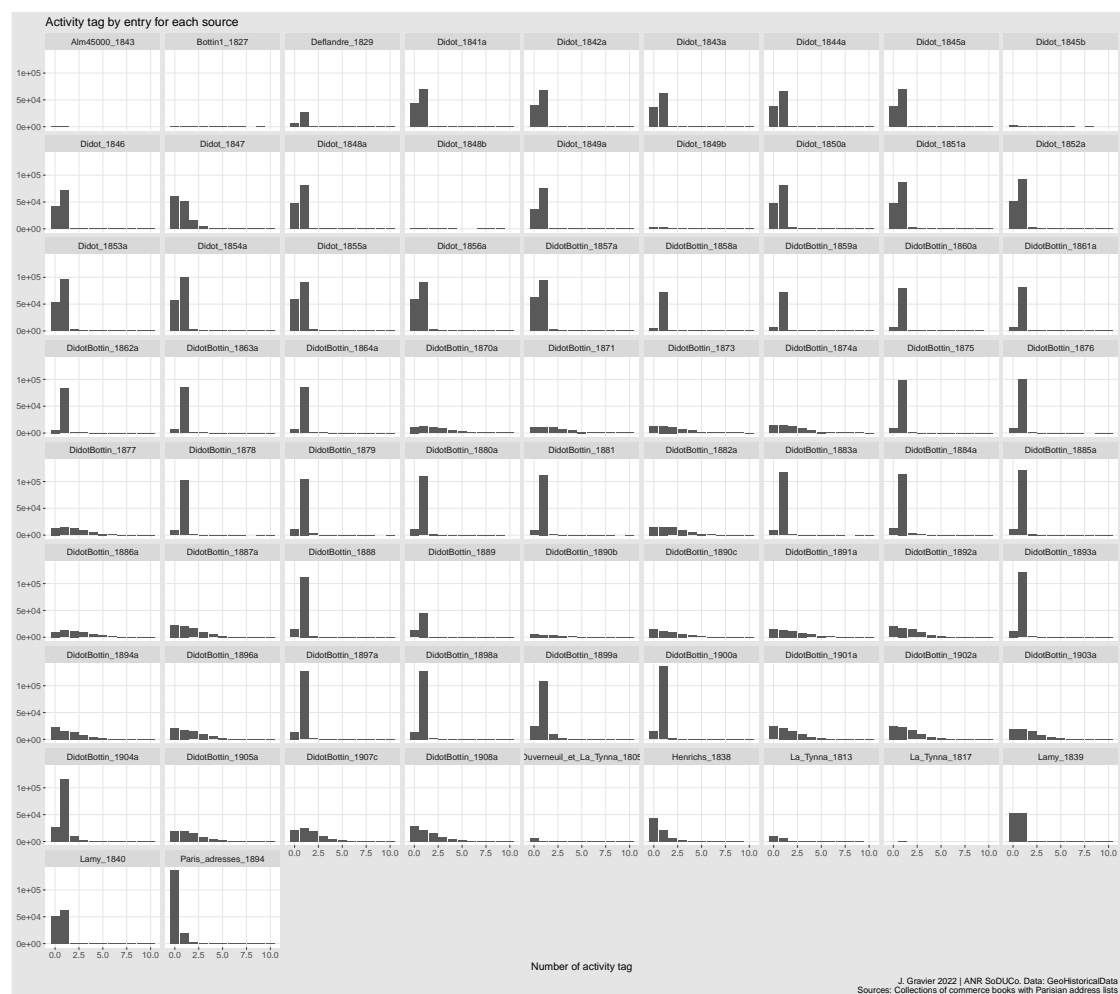


FIGURE 3 – Nombre de tags d’activités par entrée

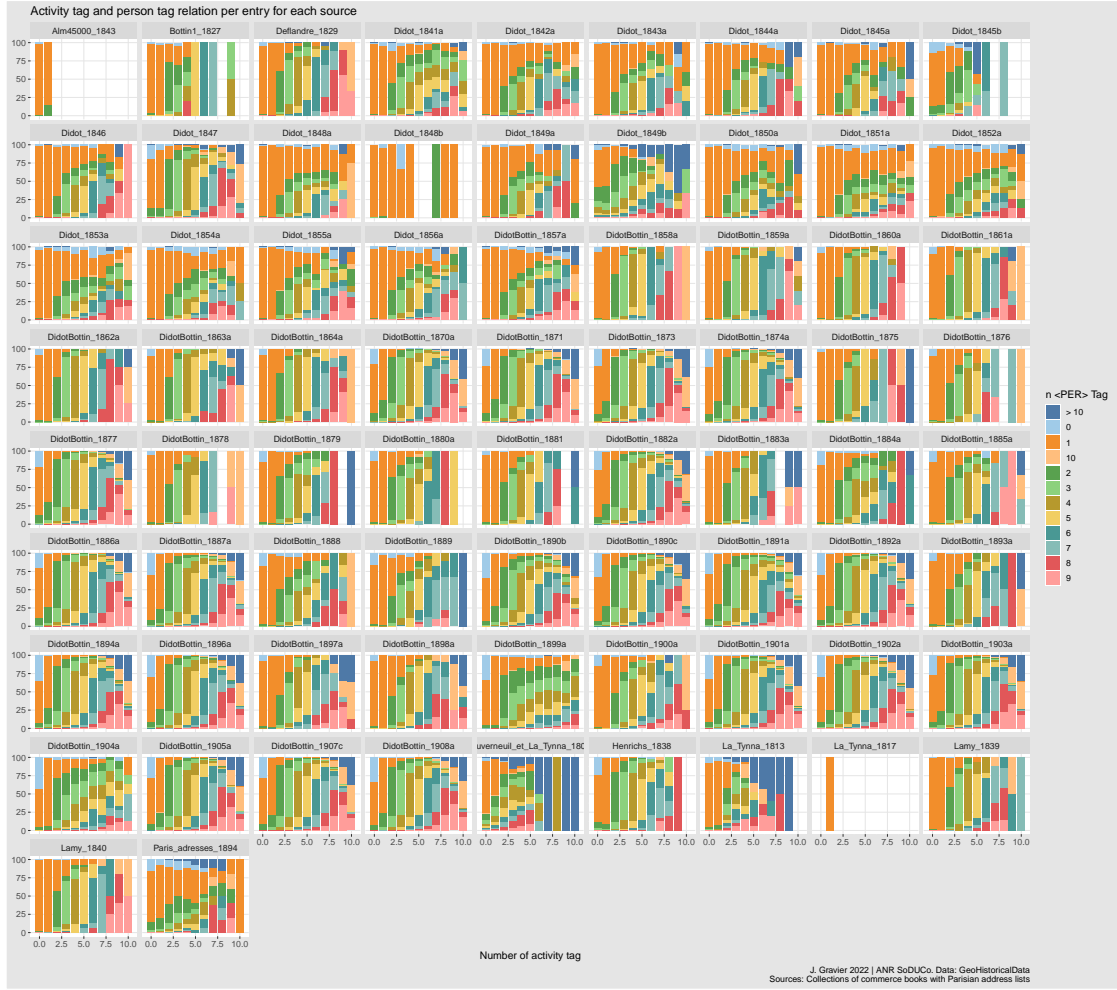


FIGURE 4 – Lien entre tags d’activités et tags de personnes par entrée

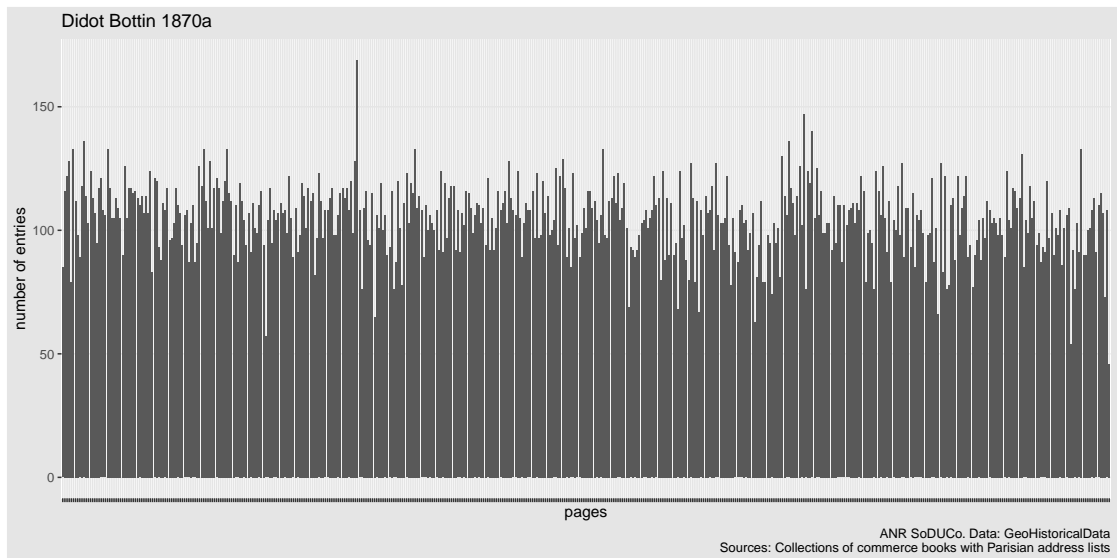


FIGURE 5 – Entrées du Bottin-Didot 1870a

6	7	numero_page	id	origin	parent	text_ocr
8		185	267	computer	260	BOC Boclet (Mue), maîtresse de langues, d'histoire et de littérature, Jacob, 1. Boclet, vins en gros, St-Dominique, 100. Boclet, produits chimiques, Bug
9		185	268	computer	260	DODEL, vins en gros, St-Dominique, 100. Boclet, produits chimiques, Bug, 9. Boclet, chaussures, bou, Richard-Lenoir, 87. Boclet, bottier, Auteuil, 10. Bocletskoy (crosse de), Châteaubriant, 19. Bocletskoy, fabricant-lampiste, Cherche- Poteskoy.
10		185	269	computer	260	Midi, 43. Bocage, cabinet de lecture, rue du Boulevard- 260 Baignolles, 17.
11		185	270	computer	260	ol.), Cots-cravates, Montmartre, 195. Bocage, librairie, avenue d'Orléans, 20. Bocage (Mie), lingère, bou, Malesherbes, 73.
12		185	271	computer	260	pocall. Bocahut (L.), boulanger, avenue d'Italie, 10. Bocahut, coloniste, Champollin, 15. Bocahut (L.), limonadier, Dunkerque, 79, et
13		185	272	computer	260	
14		185	273	computer	260	

FIGURE 6 – Exemple d'entrées multiples du Bottin-Didot 1870a

```

{
  "id": 282,
  "origin": "computer",
  "box": [
    232,
    362,
    512,
    52
  ],
  "type": "ENTRY",
  "parent": 262,
  "text_ocr": "Arnaud-Jeanti, O. *5s, maire du 3e arrondis-\nsement, Quatre-Fils, 5.",
  "addresses": [
    {
      "street_name": "Quatre-Fils",
      "street_numbers": [
        "5"
      ]
    }
  ],
  "persons": [
    "Arnaud-Jeanti"
  ],
  "ner_xml": "<PER>Arnaud-Jeanti</PER>, <TITRE>O. *5s</TITRE>, <ACT>maire du 3e arrondis-\nsement</ACT>, <LOC>Quatre-Fils</LOC>, <CARDINAL>5</CARDINAL>.",
  "geocoded": [
    {
      "full_address": "5 rue des quatre fils",
      "gazetteer.match": "rue des quatre fils",
      "gazetteer.score": 0.92918247,
      "gazetteer.source": "Denominations-emprises-voies-actuelles",
      "coarse.match": "rue des quatre fils",
      "coarse.geo.query": "rue des quatre fils",
      "coarse.geo.response": "Rue des Quatre Fils, Paris",
      "coarse.geo.score": 102.712546701,
      "coarse.geo.source": "jacoubet-paris",
      "coarse.geom": "MULTILINESTRING ((653021.5899174772202969 6862461.0686928639188409, 653076.5537840874167159 6862416.4872523285448551, 653093.3915860622655600 6862401.9707369459792972))",
      "precise.match": "rue des quatre fils",
      "precise.geo.query": "5 rue des quatre fils",
      "precise.geo.response": "5 Rue des Quatre Fils, PARIS",
      "precise.geo.score": 18.9033422408,
      "precise.geo.source": "osm-paris",
      "precise.geom": "POINT (653043.4351714409422129 6862436.5290081808343530)"
    }
  ]
}

```

FIGURE 7 – Didot-Bottin 1864, p. 102