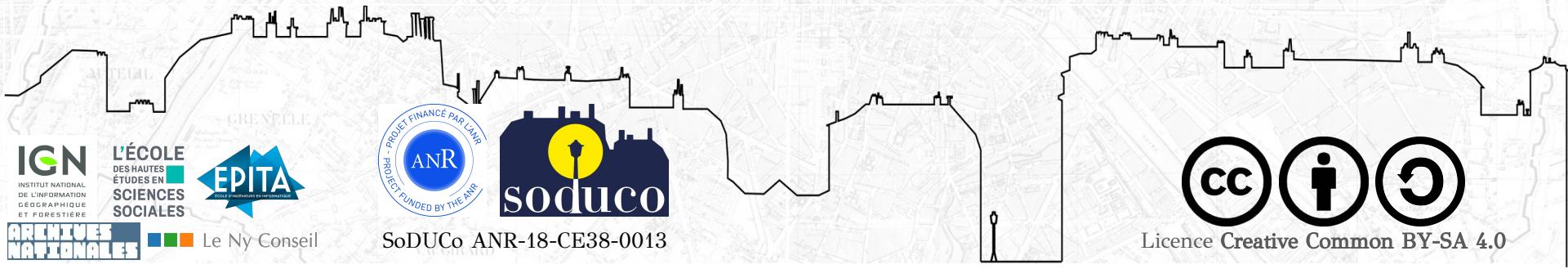


RES(T)ITUER LES ANNUAIRES COMMERCIAUX ET LES ÉVOLUTIONS DE L'ESPACE PARISIEN DU XIX^E SIÈCLE

LE PROGRAMME DE RECHERCHE SoDUCo. PANORAMA

SoDUCo Team : N. Abadie, C. Bernard, M. Hersent, J. Perret, J. Rimbault, S. Tual (LASTIG, Univ Gustave Eiffel, ENSG/IGN), E. Carlinet, J. Chazalon (LRE, EPITA), S. Baciocchi, M. Barthelemy, P. Cristofoli, B. Duménieu, J. Gravier, M. Fernandez (CRH et CAMS, EHESS), V. Weiss (CTH, Archives Nationales) et P.A Le Ny (LeNy Conseil).

4^e séance de l'Atelier SoDUCo-BnF - 6-7 novembre 2023 - Bibliothèque nationale de France, Site François Mitterrand, Paris



Projet ANR SoDUCo

(Social Dynamics in Urban Context)

Open tools, models, and data - Paris and its suburbs, 1789-1950

<https://soduco.geohistoricaldata.org/>

<https://anr.fr/Projet-ANR-18-CE38-0013>

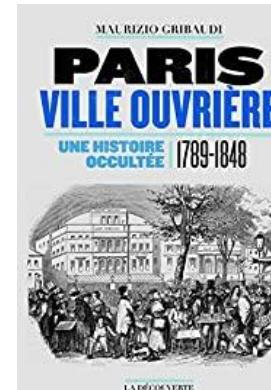
2019-2023



SOCIAL DYNAMICS IN URBAN CONTEXT (SoDUCo)

- développement de **méthodes** et de **modèles** pour l'étude de l'évolution de la structure spatiale urbaine en relation avec les activités professionnelles de la population
- reconstitution de l'évolution de la ville de Paris à l'aide de deux jeux de sources spécifiques :
 - les **cartes et plans de la ville** et
 - les **annuaires commerciaux**
- produire des **outils libres** permettant d'exploiter ces sources et **diffuser de manière ouverte** les données extraites
- se fonde sur plusieurs années de travaux initiés par **M. Gribaudi** (EHESS) et des collaborations préalablement établies entre les différents partenaires (groupe *Geohistoricaldata*, projet *Belle époque*, ...), sur la vectorisation de la *Carte de Cassini* ou l'étude de l'espace Parisien au XIX^e

M. Barthelemy & al., “Self-organization versus top-down planning in the evolution of a city”. Sci Rep 3, 2153 (2013). M. Gribaudi, *Paris, ville ouvrière. Une histoire occultée (1789-1848)*, Paris, La Découverte, 2014, 445 p. – J. Perret & al., 2015, « Roads and cities of 18th century France », Scientific Data, 15 septembre 2015, vol. 2, p. 150048. Baciocchi & al., *Engraved Footprints from the Past. Retrieving cartographic geohistorical data from the Cassini Carte de France, 1750-1789*, 2019, International Cartographic Conference 2019, Jul 2019, Tokyo, Japan. R. Cura, B. Dumenieu & al., “Historical collaborative geocoding. ISPRS International Journal of Geo-Information”, MDPI, 2018, 7 (7), pp.262.



L'ATELIER BNF-SoDUCo 2022-2023

- diffuser et **mettre en question** le travail réalisé dans le cadre de SoDUCo
- faire **dialoguer** chercheurs et spécialistes de la conservation des documents sur des objets et projets communs allant de l'exploitation des fonds à la production et diffusion de données ouvertes

Membres de l'équipe ayant participé à l'organisation des séances de l'Atelier: A. Aniesa, M. Carlin, L.-A. Charles, P. Chevallier, J. Garel-Grislin, C. Ion, O. Loiseaux, J.-D. Mellot, J.-P. Moreux, E. Netchine (BnF), N. Abadie, C. Bernard, M. Hersent, J. Perret, J. Raimbault, S. Tual (IGN-LASTIG), E. Carlinet, J. Chazalon (EPITA), **S. Baciocchi**, M. Barthelemy, P. Cristofoli, B. Duménieu, J. Gravier, M. Fernandez (EHESS, CRH et CAMS), V. Rebolledo-Dhuin (UPEC) et P.A Le Ny (LeNy Conseil).



THÉMATIQUES ABORDÉES LORS DES QUATRE SÉANCES DE L'ATELIERS :

1. *Paris, capitale géodésique. Atlas et plans à grande échelle: 1780 – 1950* (10/04/22) : **Le plan de Verniquet et l'exploitation automatisée des plans anciens.**
2. *Re(s)tituer les adresses des almanachs et annuaires commerciaux* (10/11/22) : **La chaîne de traitement et la description du contenu des annuaires.**
3. *Le circuit de traitements des documents cartographiques des XVIII^e et XIX^e siècles à la BnF* (23/05/2023, matinée) et *Analyses spatiales formalisées et écologie urbaine* (23/05/2023, après-midi)
4. *Res(t)ituer les annuaires commerciaux et les évolutions de l'espace Parisien du XIX^e siècle.*
Partager l'expérience et les ressources du programme ANR SoDUCo (6 et 7 Novembre 2023) : Science ouverte - Métadonnées, catalogage & diffusion de données - Extraction & structuration de données - Données historiques numériques – Evaluation

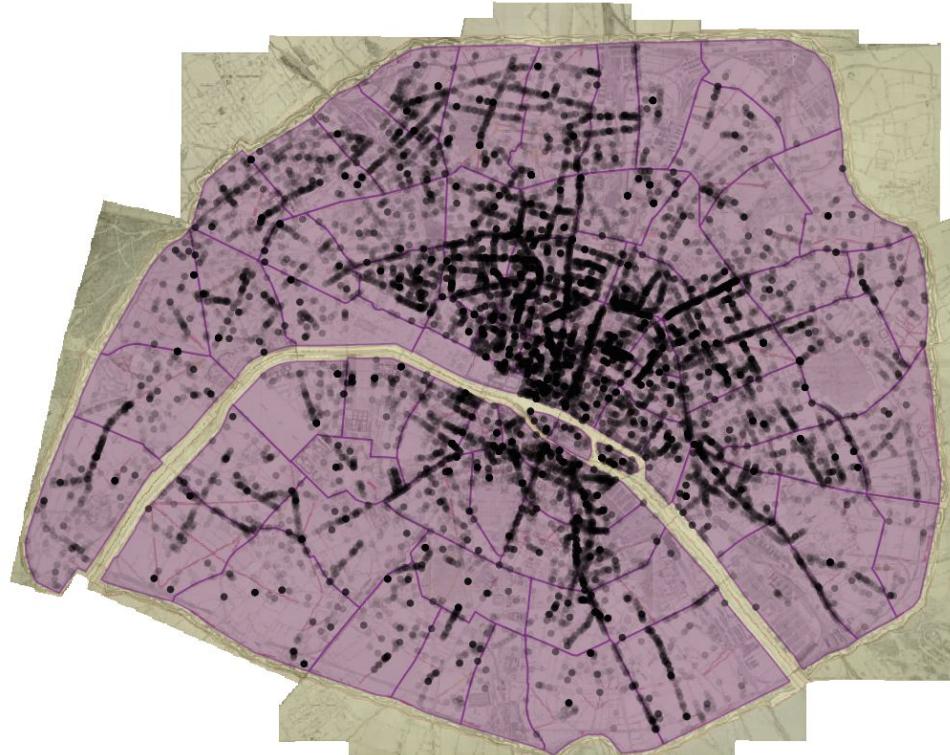


Un cadre global

Réutilisation d'un jeu de données

- sources primaires utilisés?
- traitements utilisés ?
- hypothèses effectuées ?
- Comment
 - reproduire les traitements ?
 - répliquer les traitements ?
 - modifier une étape ?

(hypothèse différente)



Le cadre global du projet SoDUCo

- **Science ouverte**
 - Traçabilité et reproductibilité des traitements et données jusqu'aux sources primaires
 - Explication des hypothèses implicites
 - Réflexion sur les formats de diffusion des données et métadonnées
 - Tout est ouvert (ou le sera bientôt)
- **Interrogation critique** (réflexive) des sources, des traitements et des données
 - Évaluation de la qualité des données produites à chaque étape
 - Évaluation de la qualité des sources
 - Question du niveau / de la granularité d'analyse (“en détail et en quantité”)
 - Utilisation des données brutes au niveau de détail de la source où elles sont extraites
- **Cumulativité**
 - Pérennisation des données
 - Pérennisation des outils
 - Pérennisation du collectif de travail
 - Pérennisation pour les institutions patrimoniales

3 questions



- Quelles sources utiliser ?
- Comment extraire des informations dans les sources ?
- Comment utiliser, interpréter et analyser les données issues des sources ?

Quelles sources utiliser ?

Des possibilités

Gallica-BnF

Archives Nationales

Retro-news

Bibliothèques spécialisées (BHdV, BHVP, etc.)

Bibliothèque de Stanford

Quelques questions

- élaborer à partir de sources éparses des **corpus de documents** adaptés aux **questions de recherche** ?
 - conserver la **trace des divers traitements** qui leur sont appliqués ainsi que les **états intermédiaires et finaux** des données numériques qui en sont extraites ?
 - créer les conditions d'une **cumulativité** à la fois du point de vue de la **recherche** et du point de vue des **institutions patrimoniales** ?

Quels **principes**, quels **outils** et quel **modèle de travail** adopter pour aller plus loin ensemble sur d'autres sujets/objets et avec d'autres acteurs/publics ?



{BnF
Gallica



David Rumsey Map Collection
CARTOGRAPHY ASSOCIATES POWERED BY LUNA IMAGING



SESSION MÉTADONNÉES, CATALOGAGE & DIFFUSION DE DONNÉES

Lundi 6 novembre 2023 (9h30-12h15) en salle 70

10h00. **SoDUCo team**, Catalogage et diffusion des données dans le programme.

10h30. **Questions**

10h45. **J. Schoonman, B. Spaan et M. Claeys Bouuaert (Allmaps)**, Présentation du consortium Allmaps

11h05. **N. Vigneron (Wikimedia France)**, Science participative, science ouverte. L'expérience de Wikimedia France

11h25. **A. Joffres (Huma-num)**, Données, outils et communautés pour recherche en SHS

11h45. **Questions**

12h00. **Discussion – Table ronde**

Comment extraire des informations dans les sources ?

Extraction des bâtiments et des îlots dans les cartes historiques

Thèse de Yizi Chen (2020-2023)

Modern vectorization and alignment of historical maps: An application to Paris Atlas (1789-1950):

<https://hal.science/tel-04106107v1>

Vectorisation de cartes historiques

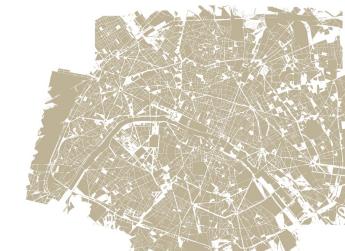
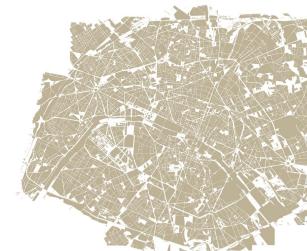
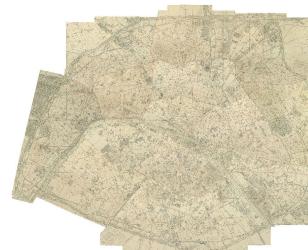
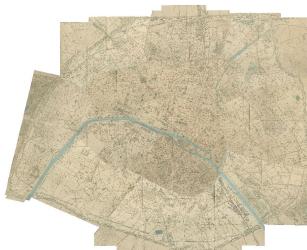
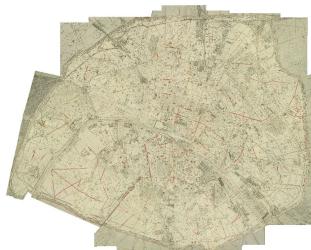
Atlas municipal des vingt arrondissements de Paris. 1925. Bibliothèque de l'Hôtel de Ville. Ville de Paris.

<http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935524>



Haut : Extrait de l'*Atlas municipal des vingt arrondissements de Paris. 1925. Bibliothèque de l'Hôtel de Ville. Ville de Paris* ;
bas : bâtiments et îlots morphologiques extraits.

Une application pour la construction d'îlotiers (en cours)



1878

1886

1900

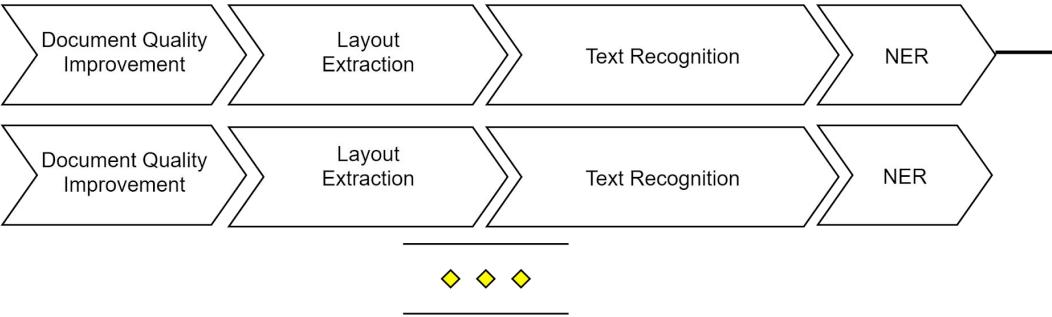
1925

1937

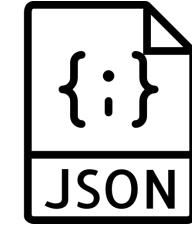
Atlas municipal des vingt arrondissements de Paris. Bibliothèque de l'Hôtel de Ville. Ville de Paris.

Chaîne de traitement des annuaires

Vue PDF 701



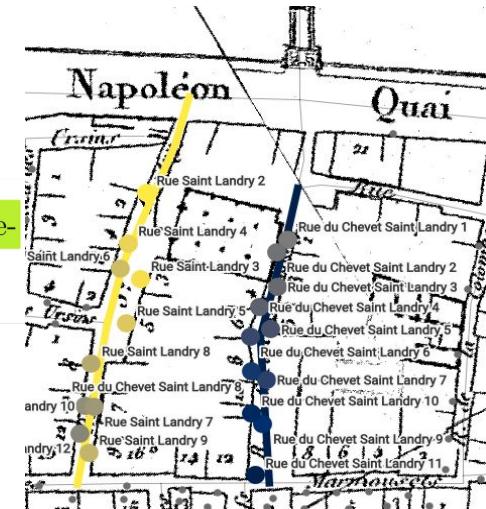
un JSON par vue



V2 data (2022-06): ~10 M d'entrées



PERSON
Dartigny (Mme), professeur de pianos, Seine-
St-Germain, 79.
ACTIVITY
LOCATION CARDINAL



SESSION EXTRACTION & STRUCTURATION DE DONNÉES

Lundi 6 novembre 2023 (15h00-17h30) en salle 70

15h00. **SoDUCo team**, Extraction et structuration de données dans le programme SoDUCo

15h30. **C. Duvette et P. Kervegan** (DER - INHA) : Chapeliers, architectes, nouveautés : géo-analyse des activités visibles et invisibles d'un quartier

15h50. **Questions.**

16h20. **W. Charles** (IRIT, Université de Toulouse Jean Jaurès) et **S. Poublanc** (Framespa, Université de Toulouse Jean Jaurès) : Représenter les territoires de la France moderne dans des graphes de connaissances : le projet OBARDI

16h40. **C. Kermorvant** (TEKLIA) : Nouvelles approches pour l'extraction d'informations dans des documents numérisés

17h00. **Questions.**

LES ATELIERS-FORMATION SODUCO – BNF DATALAB

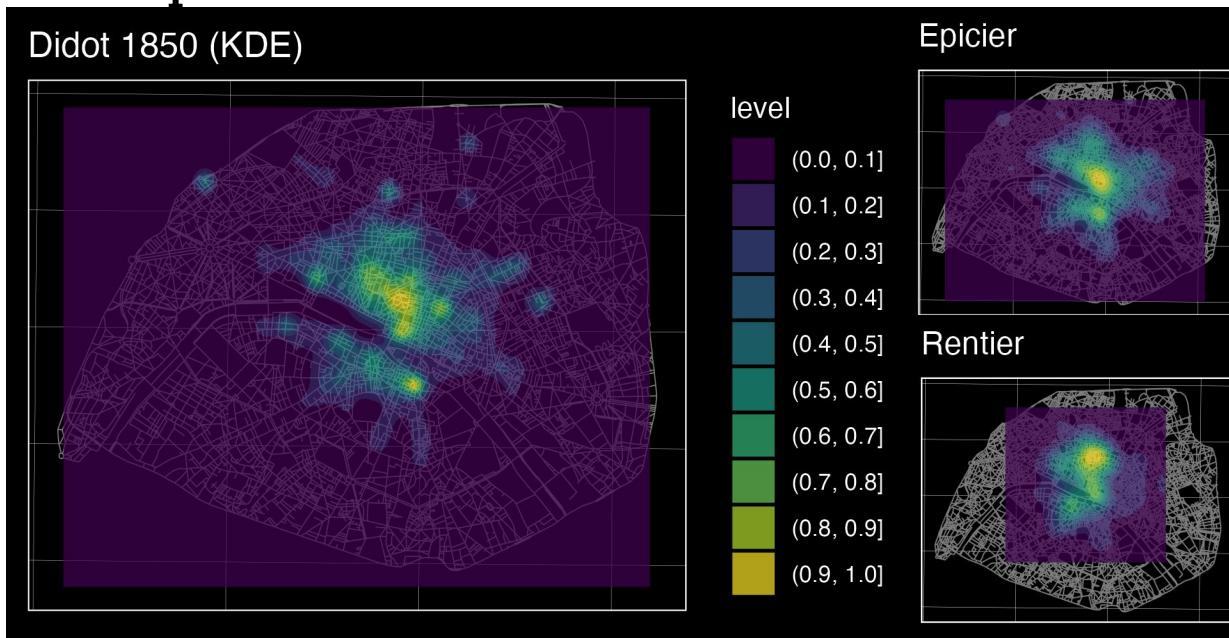
Mardi 7 novembre 2023 (13h30-17h30). Au datalab

13h30. **Atelier n°2** – Utiliser la chaîne SoDUCo d'extraction et d'enrichissement de documents textuels numériques.

13h30. **Atelier n°3** – Graphes de Connaissance Géohistoriques Professionnels.

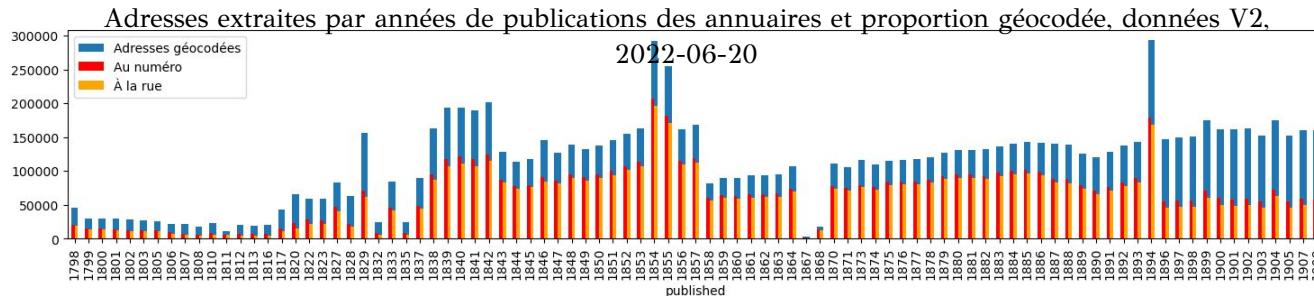
Comment utiliser, interpréter et analyser les données issues des sources ?

Des données géohistoriques



données V2, 2022-06-20 et données V2 revues

<https://doi.org/10.5281/zenodo.8388102>



14h00. Keynote - **M. Barthelemy** (CEA) et **J. Gravier** (EHESS) : « Lois d'échelle » et dynamiques des activités urbaines dans une ville en croissance.

SESSION DONNÉES HISTORIQUES NUMÉRIQUES ÉVALUATIONS CRITIQUES ET PERSPECTIVES D'UTILISATIONS

Mardi 7 novembre 2023 (09h30-12h00) en salle 70

09h30. **SoDUCo team**, Les données géohistoriques produites au cours du programme SoDUCo. Evaluation d'une « source numérique ».

10h00. **L. Kesztenbaum** (INED) et **G. Postel-Vinay** (PSE), Construction de bases de données historiques: retours d'expérience.

10h30. Questions

11h00. **C. Lemercier** (CNRS, Sciences Po-Paris CSO), Histoire quantitative, données ouvertes en SHS.

11h30. Questions et discussion finale

LES ATELIERS-FORMATION SODUCO – BNF DATALAB

Mardi 7 novembre 2023 (13h30-17h30). Au datalab

13h30. **Atelier n°1** – Exploration des données et cartographie grand public

13h30. **Atelier n°2** – Utiliser la chaîne SoDUCo d'extraction et d'enrichissement de documents textuels numériques.

13h30. **Atelier n°3** – Graphes de Connaissance Géohistoriques Professionnels.

16h45. Bilan collectif et discussion des ateliers (Mezzanine du DataLab)

Et après ?

Des suites

- **Mezanno** – plan quadriennal BnF
 - EPITA-LRE, EHESS-CRH, IGN-ENSG-LASTIG
- **Decidon** : projet soumis à l'ANR par EPITA-LRE (Marie Puren)
 - EPITA-LRE, INRIA-Almanach, EHESS-CRH, ENC-Jean-Mabillon, LARAH, BnF-DEP
- Thèses
 - **Solenn Tual** (2022-) - *Reconstitution automatique de la généalogie des parcelles et des bâtiments à partir de sources historiques.* (IGN-ENSG-LASTIG / EPITA-LRE / EHESS-CRH)
 - **Charly Bernard** (2022-) - *De sources multiples fortement hétérogènes à un référentiel d'adresses ouvert.* (IGN-ENSG-LASTIG / EHESS-CRH)

Ce séminaire est aussi

- L'occasion de faire un **retour critique** sur le projet
 - C'est vous qui allez nous aider à faire le bilan critique du projet
- De confronter nos **expériences** respectives
- De réfléchir à comment **faire mieux** par la suite
- Proposer un **cadre de discussion** formel et **informel**

Jeux de données, les outils, les articles et les plateformes

Jeux de données :

[Collection SoDUCo sur NAKALA](#)

[Annuaires historiques parisiens, 1798-1914, V2-juin 2022](#)

[A Dataset of French Trade Directories from the 19th Century \(FTD\)](#)

[A Dataset of French Trade Directories from the 19th Century for Nested NER task](#)

[NER models from "A Benchmark of Named Entity Recognition Approaches in Historical\[...\]"](#)

[Quartiers de Paris \(1860-1919\)](#)

[Populations des quartiers de Paris \(1801-1911\)](#)

[Dépôts des plans géoréférencés](#)

[ICDAR 2021 Competition on Historical Map Segmentation — Dataset](#)

Codes :

[Détection de segments](#)

[All the material \(paper, code, dataset, results\) of our DAS 2022 paper \(OCR+NER benchmark\)](#)

[All the material \(code, dataset, results\) of our Benchmark of Nested NER approaches](#)

[Démonstrateurs sur le cas des photographes parisiens](#)

[Démonstrateur v2 comprenant plusieurs datasets mis en place lors de l'atelier de liage](#)

[Segmentation et vectorisation de cartes](#)

[Des annuaires en annotations IIIF - Des cartes géoréférencées en annotations IIIF](#)

Plateformes ou sites :

[Social Dynamics in Urban Context](#)

[Catalogue - Sparql Endpoint du catalogue](#)

[Application de visualisation \(en cours\)](#)

[Viewer IIIF des annuaires traités](#)

[API REST annuaires](#)

[Datahub annuaires & atlas Ontop SPARQL endpoint for directories and professions](#)

Articles, poster & thèse:

[A Benchmark of Nested Named Entity Recognition Approaches in Historical Structured Documents](#)

[A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19th Century French Directories](#)

[\[Poster\] A Benchmark of Nested Named Entity Recognition Approaches in Historical Structured Documents - Archive ouverte HAL](#)

[Entry Separation using a Mixed Visual and Textual Language Model: Application to 19th century French Trade Directories](#)

[Création d'un graphe de connaissances géohistorique à partir d'annuaires du commerce parisien du 19ème siècle: application aux métiers de la photographie](#)

[Vectorization of historical maps using deep edge filtering and closed shape extraction - Archive ouverte HAL](#)

[Combining Deep Learning and Mathematical Morphology for Historical Map Segmentation - Archive ouverte HAL](#)

[ICDAR 2021 Competition on Historical Map Segmentation - Archive ouverte HAL](#)

[Modern vectorization and alignment of historical maps: An application to Paris Atlas \(1789-1950\)](#)

[A hidden Markov model for matching spatial networks](#)

[Modèles et outils pour la publication de métadonnées d'archives géographiques et de leurs données dérivées](#)

[Regard de... La géographie théorique et quantitative pour étudier les usages du concept de quartier en archéologie](#)

[Evolution of the urban structure of Paris during the 19th century. Building a temporal typology of urban activity categories](#)

[Exploitation des données Soduco : premiers éléments de l'évolution des activités parisiennes au fil du siècle](#)

RES(T)ITUER LES ANNUAIRES COMMERCIAUX ET LES ÉVOLUTIONS DE L'ESPACE PARISIEN DU XIX^E SIÈCLE

Lundi 6 novembre 2023 (9h30-17h30) en salle 70 – Buffet repas au Belvédère (12h15-13h30)

09h30. M. Carlin (BnF DataLab) : *Ouverture de la journée & SoDUCo team, Le programme de recherche SoDUCo. Bilan*

10h00. SESSION « MÉTADONNÉES, CATALOGAGE & DIFFUSION DE DONNÉES » avec J. Schoonman, B. Spaan et M. Claeys Bouuaert (Allmaps), N. Vigneron (Wikimedia France) et A. Joffres (Huma-num)

14h00. KEYNOTE - M. Barthelemy (CEA) et J. Gravier (EHESS) : « Lois d'échelle » et dynamiques des activités urbaines dans une ville en croissance.

15h00. SESSION « EXTRACTION & STRUCTURATION DE DONNÉES » avec C. Duvette, P. Kervegan (INHA), W. Charles, S. Poublanc (Univ. Toulouse J. Jaurès) et C. Kermorvant (TEKLIA)

Mardi 7 novembre 2023 (9h30-12h00) en salle 70 – Buffet repas au Belvédère (12h15-13h15)

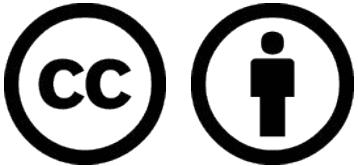
9h30. SESSION « DONNÉES HISTORIQUES NUMÉRIQUES. ÉVALUATIONS CRITIQUES ET PERSPECTIVES D'UTILISATIONS » avec L. Kesztenbaum (INED), G. Postel-Vinay (PSE), C. Lemercier (CNRS).

Mardi 7 novembre 2023 (13h30-17h30) au BnF DataLab

13h30. ATELIERS-FORMATIONS SoDUCo – BNF DATA LAB (sur inscription, places limitées)

(Ré)utiliser cette présentation

Licence Creative Common BY 4.0



N'hésitez pas à **réutiliser** ces contenus et nous **citer**.

*Sauf citation ou mention contraire,
nous sommes les auteurs des contenus.*

Consultez le diaporama pendant et après la présentation

<https://link.infini.fr/soduco>

Les autres présentations de l'atelier :

<https://link.infini.fr/soduco-catalog>

<https://link.infini.fr/soduco-lois-dechelle>

<https://link.infini.fr/soduco-extraction>

<https://link.infini.fr/soduco-geohistorique>

