



# Moving Forward in Administrative History: Encoding the Département de La Seine and Paris Yearbooks (1883–1970)

CAROLE LAMOUREUX 

ELSA CAMUS

\*Author affiliations can be found in the back matter of this article

COLLECTION:  
ARCHIVAL CITY

RESEARCH PAPER



## ABSTRACT

The administrative yearbooks of the Seine departments and of the city of Paris are crucial sources for understanding the organization of these institutions, and also for finding one's way around the archives related to the administration of the city. The project carried out as part of the Archival City programme aimed to facilitate access to these yearbooks. Digitalized by the Archives de Paris, which hold them, these documents have been transcribed and encoded in XML-TEI format. Using this fundamental basis, data exports have been performed in handy formats and a visualisation website is currently under construction (planned for the beginning of 2024). Our paper intends to present this project, the sources we have been working on and the results we have obtained. It especially aims to enable the scientific community to use, reuse and enhance the resultant datasets.

## CORRESPONDING AUTHOR: Carole Lamoureux

Equipe d'accueil Analyse  
comparée des pouvoirs (ACP),  
Université Gustave Eiffel,  
Champs-sur-Marne, France

[carolelamoureux.l@gmail.com](mailto:carolelamoureux.l@gmail.com)

## KEYWORDS:

administrative history;  
urban history; encoding;  
digitalization; yearbooks; Paris  
(France)

## TO CITE THIS ARTICLE:

Lamoureux, C., & Camus, E.  
(2024). Moving Forward  
in Administrative History:  
Encoding the Département de  
La Seine and Paris Yearbooks  
(1883–1970). *Journal of  
Open Humanities Data*, 10:  
18, pp. 1–15. DOI: [https://doi.  
org/10.5334/johd.186](https://doi.org/10.5334/johd.186)

The digitization, transcription, encoding and visualisation of the administrative yearbooks of the Département de la Seine (Seine department) and the Ville de Paris (city of Paris, France) held by the Archives de Paris<sup>1</sup> has been one of the main projects of the “Greater Paris” study of the Archival City programme.<sup>2</sup> Under the scientific direction of Loïc Vadelorge (Professor, Université Gustave Eiffel), it began in 2019, with the end of 2023 as the deadline. It aims to facilitate access to those documents and their contents, for the period 1883–1970. The Archives de Paris, the Bibliothèque de l’Hôtel de Ville<sup>3</sup> and the École des Ponts ParisTech<sup>4</sup> have been crucial partners in this project, both scientifically and logistically.

The purpose of this paper intends is to introduce the scientific potentials of the administrative yearbooks of the Seine department and of the city of Paris, our project, its data and results, whose main objectives are: consultation and reuse by the scientific community. It is linked to the papers written by Loïc Vadelorge (2024) and Paul Lesieur (2023) on the scientific stakes of editing these yearbooks and on extracting specific information.

## (1) CONTEXT AND MOTIVATION

### (1.1) STATE OF APPLICATION OF THE ADMINISTRATIVE YEARBOOKS

The administrative yearbooks of the Seine department and of the city of Paris are a series of printed documents beginning in 1883. The corpus we worked on ends in 1970, a date chosen for scientific purposes which are explained later.<sup>5</sup> Each edition gives a list of the administrative departments, with details on their functions and assignments, their composition (directors, secretaries, engineers, employees, etc.), locations and other practical information. Most of them are entitled “Personnel et attributions des services du département de la Seine et de la Ville de Paris” (1885–1911) or “Organisation et attributions des services du département de la Seine et de la Ville de Paris” (1920–1966, except for 1955 and 1957). Without counting the separate *addenda*, the 39 documents are irregularly ordered through time: 1883, every year from 1885 to 1887 (1885–1887), 1889, 1891–1894, 1896, 1897–1898, 1903–1905, 1907, 1911, 1920, 1925–1926, 1930, 1932, 1935, 1941, January 1942<sup>6</sup>, July 1942, 1945, 1948–1950, 1953, 1955, 1957, 1959, 1961–1962, 1966, 1968, 1970.

Before the beginning of our project, the Archives de Paris had neither digitalized these yearbooks nor promoted them in any way other than in encouraging researchers to consult them. As explained by the researchers of Archival City’s team (Vadelorge, 2024; Lesieur, 2023), these documents are well known by the historians working on the area concerned. They are used to find the specific information they are looking for, but, before our project, they had no complete idea of how this administration was organized and how it evolved.

Similarly, the archivists of the Archives de Paris used the yearbooks when searching for specific departments, functions or individuals, in order to classify their fonds and guide historians. However, they have not yet had the time and opportunity to work on the complete history of the administrative departments which are the creators of the archives they are holding. Due to the complexity of this administration, its genealogy and because of the lack of complete reference material about them, it can be quite difficult for researchers to find their way around the fonds.

### (1.2) SCIENTIFIC GOALS

Our main goal has been to make the yearbooks accessible online and to facilitate the work of researchers who do not have to go through every page of every document immediately. Thus, we originally aimed to transcribe these documents, after their digitalization by the Archives de Paris. The next step was to encode them in order to characterize every piece of information

---

1 Website of the Archives de Paris: <http://www.archives.paris.fr> (last accessed: 19 November 2023).

2 Archival City, 2019–2023.

3 Website of the Bibliothèque de l’Hôtel de Ville: <https://www.paris.fr/lieux/bibliotheque-de-l-hotel-de-ville-bhdv-17> (last accessed: 19 November 2023).

4 École de Ponts ParisTech, “Bibliothèque”: <https://ecoledesponts.fr/documentation> (last accessed: 19 November 2023).

5 Reference codes of our corpus, 1883–1970: Archives de Paris, PER232 1–28.

6 Here, the months are only specified for the two yearbooks which have been edited in the same year.

and enable both the search and extraction of specific data, such as references to places and individuals. From the very beginning, one of the principal objectives of this encoding has been to highlight and technically structure the organization of the administration, thus enabling the visualisation of the departments in the form of organigrams. However, the yearbooks were not conceived and elaborated as real organigrams but as textual presentations, with title levels intended to make information clearer. For this reason, in what follows, we will refer to “hierarchies” instead of “organigrams”.

As the corpus is large, it was obvious that not all the information of all the yearbooks could be worked on from 1883 to nowadays. We decided to end our corpus a few years after the reorganization of the Paris region and the dissolution of the Seine department, at the beginning of 1968. Thus, we worked on the yearbooks related to the Seine department (1883–1966) and to the very beginning of the reorganization of the Paris municipality (1968–1970). As time was limited, we have left out the *addenda*, which notify changes between two editions, in order to focus on documents that have a fairly similar structure and can be easily linked together. We have also left out the yearbook edited in January 1942, as that of July 1942 is very close in time and is more precise. Thus, we have ended up with a corpus of 38 yearbooks, which is still quite large.

Alongside those general objectives, we also wanted to answer the specific needs of the researchers. Thus, the parts related to the head of the Préfecture, to city planning, to the public roads and to archives were priorities for more precise encoding. We also thought of elaborating a thesaurus of the functions and assignments of the departments<sup>7</sup>, and to normalize and geo-reference the locations of offices. However, those two important tasks needed more resources proper completion and were set aside for later projects.

We enhanced and made accessible all the progress we could make through the elaboration of visualisation tools, such as specific data exports and a visualisation website enabling researchers to consult the yearbooks easily.

### (1.3) SIMILAR PROJECTS AND INSPIRATIONS

When it was launched in 2019, the administrative yearbooks project had no equivalent, but followed the footsteps of other initiatives carried out by various institutions, starting with the organization charts of the central administrative departments of the Ministère de l'Ecologie (from 1830 to the present day):<sup>8</sup> in addition to access to source files by year, a search engine enables detailed browsing the directories by departments and names of individuals.

The *Millefeuille* project carried out by the Archives Nationales (France) was another source of inspiration (Fekete, & Ogilvie, 2008). It was initiated in 2008 but not maintained since. Its aim was to offer a visualization of administrative structures from the 18<sup>th</sup> to the 20<sup>th</sup> centuries, as mentioned in official Almanacs, thanks to XML-TEI encoding.

For several years, several archive departments have been offering interactive visualization tools for accessing their databases of persons, communities and places. For instance, *L'Encyclo*, developed at Orléans,<sup>9</sup> and the visualization of data belonging to the services of the city of Lille<sup>10</sup> give access to archival descriptions through administrative functions.

During the first phases of our project, we discussed on a regular basis with the “Annuaire et adresses” of the Paris Time Machine workgroup.<sup>11</sup> It aimed to create a geo-historical repository of Paris, with the “Annuaire des Propriétaires de Paris” as the main source for the extraction and geo-referencing of addresses.

---

<sup>7</sup> On the importance of such a thesaurus, see Vadelorge (2024).

<sup>8</sup> Ministère de l'Ecologie, du développement durable et de l'Energie, Comité d'Histoire, Website *Les directions d'administration centrale des origines à nos jours*: <https://www.histoire-dac.developpement-durable.gouv.fr/index.xsp> (last accessed: 19 November 2023).

<sup>9</sup> Archives municipales et métropolitaines d'Orléans, *L'Encyclo*: <https://archives.orleans-metropole.fr/histoires-dorleans/lencyclo> (last accessed: 19 November 2023).

<sup>10</sup> Archives de la Métropole européenne de Lille, *Les Archives de la MEL*: [https://archives.lillemetropole.fr/data/files/mel.diffusion/images/DATAVIZ/Livrable\\_2022/Visualisation\\_mel\\_12152022.html](https://archives.lillemetropole.fr/data/files/mel.diffusion/images/DATAVIZ/Livrable_2022/Visualisation_mel_12152022.html) (last accessed: 19 November 2023).

<sup>11</sup> Since the beginning of our project, Paris Time Machine has been renamed Projets Time Machine. Projets Time Machine, “Groupe Annuaire et Adresses”: <https://ptm.huma-num.fr/chantier-groupe-annuaire-et-adresses/> (last accessed: 19 December 2023).

The open access platform *Fabrique numérique du passé*, dedicated to datasets related to urban history, also gives recent examples of works elaborated through the compilation of yearbooks, such as the “Paris Didot-Bottin année 1877” dataset of the Institut national d’histoire de l’art (INHA).<sup>12</sup>

## (2) DESCRIPTION OF DATASETS

This project has led to the creation of three datasets: the main one, containing the XML-TEI files; and two others consisting of data exports.

### (2.1) MAIN DATASET: XML-TEI ENCODINGS

#### URL

<https://doi.org/10.34847/nkl.9efe494x>

#### Object name

Archival City Greater Paris- Administrative yearbooks for the City of Paris and the Seine Department (1883–1970): XML-TEI encodings.

#### Format names and versions

XML, RNG, XSLT, CSV, Excel.

#### Creation dates

2020-04-01 – 2023-11-01.

#### Dataset creators

Elsa Camus and Carole Lamoureux (conceptualization, data curation, methodology, supervision, visualization – Université Gustave Eiffel); Gaël Donneger and Vincent Tuchais (digitalization – Archives de Paris); Cassandre Maubert (data curation – Université Gustave Eiffel); Sonya Bensaadi, Fanny Brière, Nathan Dacier-Falque, Valentin Devillepoix, Andréa Dorin, Jeanne Fras, Anahi Haedo and Léna Humbert (data curation – interns of the Université Gustave Eiffel).

#### Language

French

#### License

Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0).

#### Repository name

Nakala.

#### Publication date

2023-11-14.

### (2.2) PDFS WITH HIERARCHIES AND IMAGES

#### URL

<https://doi.org/10.34847/nkl.fcd78p1m>

#### Object name

Archival City Greater Paris- Administrative yearbooks for the City of Paris and the Seine Department (1883–1970): images with hierarchies.

---

<sup>12</sup> Fabrique numérique du passé, “Paris Didot-Bottin année 1877”: [https://www.fabriquenumeriquedupasse.fr/explore/dataset/paris\\_jobs\\_with\\_tags\\_richelieu\\_project\\_bottin1877/](https://www.fabriquenumeriquedupasse.fr/explore/dataset/paris_jobs_with_tags_richelieu_project_bottin1877/) (last accessed : 19 December 2023).

## Format names and versions

PDF.

## Creation dates

2023-11-01.

## Dataset creators

Carole Lamoureux (conceptualization, data curation, methodology – Université Gustave Eiffel);  
Gaël Donneger and Vincent Tuchais (digitalization – Archives de Paris).

## Language

French

## License

Creative Commons Attribution Non Commercial Share Alike 4.0 International. (CC-BY-NC-SA-4.0).

## Repository name

Nakala.

## Publication date

2023-11-14.

## (2.3) DATA EXPORTS IN CSV AND MARKDOWN

### URL

<https://doi.org/10.34847/nkl.def5o41e>

### Object name

Archival City Greater Paris- Administrative yearbooks for the City of Paris and the Seine Department (1883–1970): data exports.

## Format names and versions

CSV, Markdown.

## Creation dates

2023-11-01.

## Dataset creators

Carole Lamoureux (conceptualization, data curation, methodology – Université Gustave Eiffel)

## Language

French

## License

Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0).

## Repository name

Nakala.

## Publication date

2023-11-14.

## (3) METHOD

### (3.1) DIGITALIZING AND TRANSCRIBING

The collection of administrative yearbooks was digitized in December 2019, i.e. at the beginning of the project. These original images, which have not been placed in open access and are held by the Archives de Paris, are organized with one file for each archival reference code and one JPG image for each double page (200 dpi for diffusion files). The first task undertaken by Elsa Camus in 2020 was the quality control of the digitization process (identifying errors such as missing pages and unusable images), followed by a structural analysis of the corpus.

The first step in exploiting the corpus was to recover the textual content, by initiating an OCR process on the images. Using Limb Processing software incorporating the Abby10 OCR Engine, installed on the Copibook digitizer made available to the project by the library of the École des Ponts ParisTech, it was possible to separate the bi-feuillets into mono-feuillets, and to obtain OCRs of the yearbooks in PDF, TXT and XML/ALTO formats. The quality of the OCRs obtained, with recognition rates below 95%, was not sufficient for text extraction and text mining.

In the second phase, we considered the use of an automatic handwritten text recognition software application (HTR) for more accuracy: OCR4All, Transkribus and eScriptorium were tested. Finally, the eScriptorium<sup>13</sup> open-source platform was chosen. This platform, developed since 2020 by the Scripta-PSL research group, is free and has an ergonomic online interface. We tried to import the original OCRs obtained with Limb and to make models manually (with a dozen pages) to make new OCRs. However, each model could take more than two weeks to be made and the result was not satisfactory enough to spend the time actually taken. Therefore, we decided to correct the text formats of the original Limb OCR manually.

### (3.2) ENCODING IN XML-TEI

The choice of encoding the yearbooks in XML-TEI<sup>14</sup> format answered the wish to analyse and export the information strictly as it is displayed in the sources. Our purpose was to facilitate access and leave a more advanced study of administrative history for a later date. For instance, in other circumstances, the standard XML-EAC<sup>15</sup> format would have enabled us to encode the administrative hierarchies, by compiling the original data and thus creating a perspective with the sources.

The selection of the main XML-TEI elements used in this project and of the general schema associated with it was made by Elsa Camus. Adaptations and evolutions were made by Carole Lamoureux throughout the encoding process, in respect to the special cases encountered in each yearbook and to the suggestions of the interns working on them.

The overall schema consists of one XML document referencing the whole corpus and incorporating both the XML files produced for each yearbook and the index of persons presented later in this document.

The specific schema of the files encoding the yearbooks themselves consists of a hierarchy of nested “org” elements. The purpose of this was to have tree structures portraying the organization of the administration and the links between a department and its subdepartments, for example, both intellectually and technically. Each “org” element embeds the title or name of the structure in an “orgName” and may contain other XML elements indicating references to individuals (“persName”), addresses of offices (“placeName”), details related to the assignments of this structure (“desc[@type=’attributions’]”) or presenting it more generally (“desc[@type=’presentation’]”), for example (Figures 1 and 2).

One of the main challenges of this schema, which focuses more on the intellectual information than on the text organization, was to account for the difference between the strict title contents

---

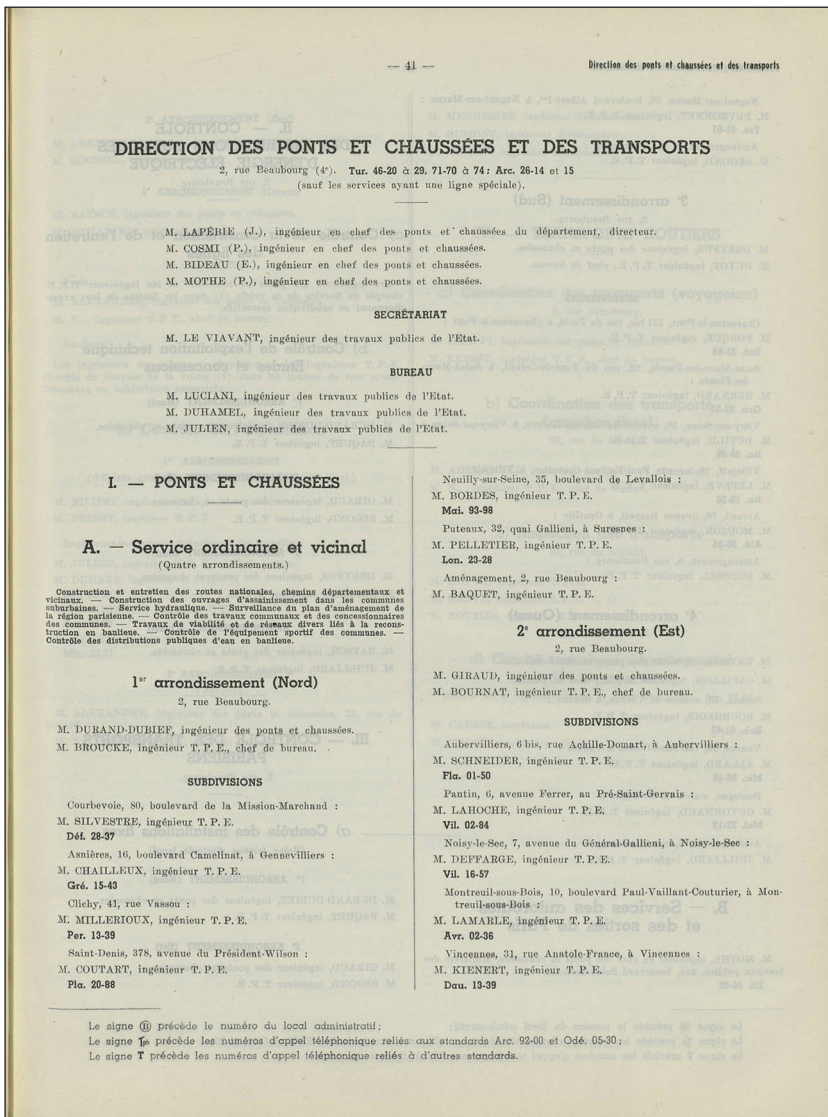
<sup>13</sup> Chagué, 2022.

<sup>14</sup> Website and documentation of the *Text Encoding Initiative (TEI)*: <https://tei-c.org/> (last accessed: 19 November 2023).

<sup>15</sup> Website and documentation of the *Encoded Archival Context for Corporate Bodies, Persons and Families (EAC-CPF)*: <https://eac.staatsbibliothek-berlin.de/> (last accessed: 19 November 2023).

and the structure names. For instance, all the yearbooks from 1885 to 1926 have a part (“org”) “3° Services et commissions se rattachant aux bureaux de la Préfecture”. It obviously consists of a title, and not of a structure name. However, we could not be a 100% sure of this distinction for many contents. Consequently, we decided to opt for a less cumbersome way of accounting for this type of distinction, and to leave what we have interpreted as strictly titles (and not structure names) in “orgName” elements with the value “0” for the attribute “@n”.

Figure 1 Edition of 1955, p.41, digitalization.



```

3221 <org n="2" role="voitrie">pb n="41" face="AD075B1_FER232_0020_0043.jpg"/>
3222 <orgName>DIRECTION DES PONTS ET CHAUSSÉES ET DES TRANSPORTS</orgName>
3223 <placeName>2, rue Beaubourg (4e). </placeName><district type="emplacement-bureau">Tur. 46-20 à 29, 71-70 à 74; Arc. 26-14 et 15
3224 (sauf les services ayant une ligne spéciale).</district>
3225 <persName ref="#lapébie">M. LAPÉBIE (J.), <roleName type="titre">ingénieur en chef des ponts et chaussées du département</roleName>,
3226 <affiliation type="position-hiérarchique">directeur</affiliation></persName>
3227 <persName ref="#cosmi">M. COSMI (P.), <roleName type="titre">ingénieur en chef des ponts et chaussées</roleName></persName>
3228 <persName ref="#bideau">M. BIDEAU (E.), <roleName type="titre">ingénieur en chef des ponts et chaussées</roleName></persName>
3229 <persName ref="#mothe">M. MOTHE (P.), <roleName type="titre">ingénieur en chef des ponts et chaussées</roleName></persName>
3230
3231 <org n="3">orgName:SECRÉTARIAT</orgName>
3232 <persName ref="#levivant">M. LE VIAVANT, ingénieur des travaux publics de l'Etat.</persName></org>
3233
3234 <org n="3">orgName:BUREAU</orgName>
3235 <persName ref="#luciani">M. LUCIANI, <roleName type="titre">ingénieur des travaux publics de l'Etat</roleName></persName>
3236 <persName ref="#duhamel">M. DUHAMEL, <roleName type="titre">ingénieur des travaux publics de l'Etat</roleName></persName>
3237 <persName ref="#julien">M. JULIEN, <roleName type="titre">ingénieur des travaux publics de l'Etat</roleName></persName></org>
3238
3239 <org n="0">orgName:surplus:I. — </surplus>PONTS ET CHAUSSÉES</orgName>
3240
3241 <org n="3">orgName:surplus:A. — </surplus>Service ordinaire et vicinal</orgName>
3242 <desc type="composition">(Quatre arrondissements.)</desc>
3243 <desc type="attributions">Construction et entretien des routes nationales, chemins départementaux et
3244 vicinaux. — Construction des ouvrages d'assainissement dans les communes
3245 suburbaines. — Service hydraulique. — Surveillance du plan d'aménagement de
3246 la région parisienne. — Contrôle des travaux communaux et des concessionnaires
3247 des communes. — Travaux de viabilité et de réseaux divers liés à la reconstruction en banlieue. — Contrôle de l'équipement sportif des communes. —
3248 Contrôle des distributions publiques d'eau en banlieue.</desc>
3249
3250 <org n="4">orgName:1er arrondissement (Nord)</orgName>
3251 <placeName>2, rue Beaubourg.</placeName>
3252 <persName ref="#duranddubief">M. DURAND-DUBIEF, <roleName type="titre">ingénieur des ponts et chaussées</roleName></persName>
3253 <persName ref="#broucke">M. BROUCKE, <roleName type="titre">ingénieur I.P.E.</roleName>, <affiliation type="position-hiérarchique">chef de bureau</affiliation></persName>
3254 <org n="0">orgName:SUBDIVISIONS</orgName>
3255
3256 <org n="5">orgName:Courbevoie</orgName><placeName>80, boulevard de la Mission-Marchand </placeName>
3257 <persName ref="#sylvestre">M. SYLVESTRE, <roleName type="titre">ingénieur T.P.E.</roleName></persName>
3258 <district type="emplacement-bureau">Def. 28-37</district></org>
3259
3260 <org n="5">orgName:Anières</orgName><placeName>16, boulevard Camelinat, à Gennevilliers </placeName>
3261 <persName ref="#chailleux">M. CHAILLEUX, <roleName type="titre">ingénieur T.P.E.</roleName></persName>
3262 <district type="emplacement-bureau">Gr. 15-43</district></org>
3263
3264 <org n="5">orgName:Clichy</orgName><placeName>41, rue Vassou </placeName>
3265 <persName ref="#millerieux">M. MILLERIEUX, <roleName type="titre">ingénieur T.P.E.</roleName></persName>
3266 <district type="emplacement-bureau">Per. 13-39</district></persName></org>

```

Figure 2 Edition of 1955, beginning of p.41, encoding.

The large extent of the corpus (38 yearbooks, with from 70 to more than 400 pages each) and the time taken by each step of the encoding led us to decide between selecting a few of them for extensive encoding and working with all of them for a more panoramic view (Figure 3). As the main purpose of this project was to follow the evolution of the administrative organization over the decades, the second option was chosen. The consequence of this was the necessity of selecting between the parts we wanted to encode more precisely and the data we could put aside in the context of the Archival City programme.

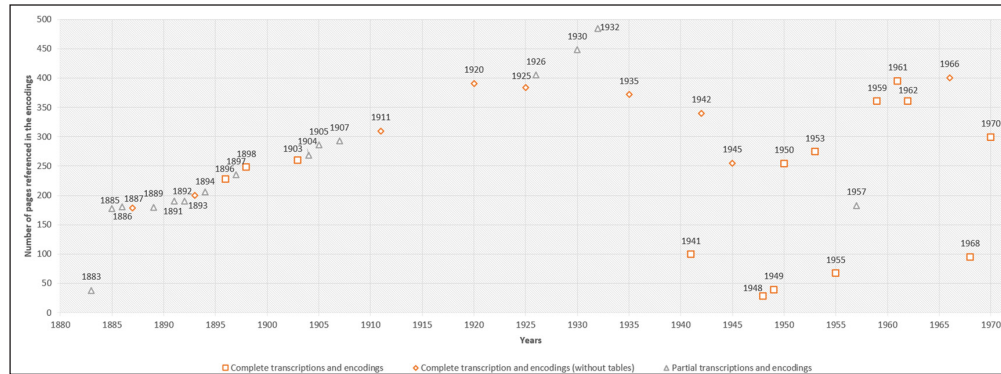


Figure 3 Yearbooks transcribed and encoded.

For this reason, the transcription has been wholly corrected and encoded for 14 yearbooks, to account for all the information that can be found in this corpus. Nine others have been corrected and encoded integrally, except for the large tables (listing physicians, for example). The 15 yearbooks left have been encoded in their structure “org”, for their titles or names of structures (“orgName”) and with the detail interesting the Archival City programme. The transcriptions which have not been corrected have been put in XML comments.

Likewise, the individuals related to our research interests were encoded more precisely, with the distinction of their roles, titles, honorific distinctions, and information about the location of their office. Moreover, with the attribute “@ref”, these references are connected to the index file which lists the individuals concerned and details their first name, last name, possible name link, and sex.

Although these encodings account more for the intellectual contents, some important markers of the layout and the materiality of the yearbooks have been indicated. Thus, the elements “pb” link the XML data with the original pages of the sources (“@n”) and with the digitalisations (“@facs”). Moreover, if a page’s degradation has rendered a part of the text unreadable, this same degradation has been encoded (“damage” elements). For most of the yearbooks, the original line breaks have been respected in the XML documents, as it makes it easier to read, but words covering two lines have been reconstituted to facilitate full-text searches.

In some cases, the copies of the yearbooks we worked on included handwritten information, like names of persons crossed out and replaced by others. As these indications may provide knowledge about changes in the administration, they have been encoded as such.

As our project focused on the data related to administrative evolutions, the introductory parts of the yearbooks (tables of contents, presentations of State administrations, etc.) and their annexes (indexes, etc.) have not been encoded. Their transcription by Limb have been put in comments in “front” and “back” elements to be worked on in later projects.

The yearbook for 1968 is a special one. It does not present the whole hierarchy, but the head of each department of the municipality of Paris, at a time when the Seine department had just been dissolved<sup>16</sup> and a complete overhaul was launched. The document structure is different from the other yearbooks and has called for the elaboration of a specific TEI schema, derived from the main one.<sup>17</sup> Moreover, the introductions of the new directors retrace their whole careers and contain a great deal of personal data about them. For this reason, an embargo

16 Décret n°67-792 du 19 septembre 1967 relatif à l’entrée en vigueur au 1<sup>er</sup> janvier 1968 de la loi du 14 juillet 1964 portant réorganisation de la région parisienne. Retrieved from <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000330666> (last accessed: 19 November 2023).

17 General schema: ACP03\_teiAnnaires.rng; schema for 1968: ACP03\_teiAnnaires\_1968.rng.



has been placed in Nakala regarding the consultation of the data of this yearbook, and the visualization website due to be launched will not include it.

The contribution of the interns has been crucial for this stage of the project, especially as many yearbooks have specific characteristics and the very time-consuming nature of the task. Each document required an average of one and a half months of work.

### **(3.3) MAKING THE YEARBOOKS AVAILABLE AND ACCESSIBLE FOR RESEARCHERS**

Throughout the encoding of the yearbooks, one of the challenges for us was to give access to our work in progress to the researchers of the team in order to make decisions and to have their feedback. The XML files were not sufficient for this purpose, as they require specific technical abilities and their sheer magnitude makes it quite hard to obtain a synthetic view without using a certain number of XPath queries. This imperative for internal dialogues allowed us to think early, and at different stages, on the best ways to make the data accessible for the purpose of making them available in open access.

Occasionally, data exports were carried out to discuss with the researchers: hierarchies in HTML, in PDF or in the form of mind maps; details of the data related to specific services, element by element, in PDF. In addition to this, the local development on eXist-db<sup>18</sup> made it possible to show the hierarchies and display the detailed information of the departments dynamically. Eventually, these discussions led to feedback and reflections on the form and the usability of these exports as much as on the informational contents and the scientific perspectives. They led to the development of the final data exports presented later in this paper.

This whole internal process clearly showed that the publication of the XML-TEI files was not sufficient. It gave hints of what could be profitable. First, the visualization website that had been conceived had to display the main scientific objects of our work and their potentialities.

Second, one of the types of exports most appreciated by the researchers of the team was the generation of one PDF for each yearbook, with the table of contents of the original document, a corresponding navigation pane and the digitalisations. Third, we had the confirmation that data exports of references to individuals and to locations would be useful for some investigations, as was the ability to display entire hierarchies with mind map software.

### **(3.4) OPENING THE YEARBOOKS TO BROADER RESEARCH PROCESSING.**

Even before starting to encode the yearbooks, we knew that we did not have the means to account for all the information of these documents and exploit them to their full potential because of their length and density. Beyond the very principles of open science, we conceived our work as a first big step in accessing and analysing the yearbooks. Thus, it became even more fundamental to dispose of an XML schema that was relatively easy to understand and use, as well as to document our files properly and in a practical way. In this perspective, the very fact of working as an evolving team (succession of interns) and of having to hand the XML files to the firm in charge of developing the visualisation website helped us to simplify the schema as much as possible and clarify our presentation of it. Indeed, at every stage, the challenge was to open the encoding of the yearbook to a new collaborator.

Alongside explanatory lists of XML elements, one of the first documentation tools used daily by the team was the RELAX NG<sup>19</sup> schema elaborated at the beginning with the Roma customization tool<sup>20</sup> and updated at each evolution of the selection of XML-TEI elements and attributes. The association of this RNG schema to the files helped especially with the filling of mandatory attributes in our XML schema: “org/@n”, “desc/@type” with a closed list of possible values, “roleName/@type”, also with a closed list of values, etc.

In the metadata of the XML files, i.e. in the “teiHeader”, the “encodingDesc” of each yearbook is filled only with information about its level of transcription and encoding. Indeed, as they

---

18 eXist-db presentation and documentation: <http://exist-db.org> (last accessed: 19 November 2023).

19 RELAX NG presentation and documentation: <https://relaxng.org/> (last accessed: 19 November 2023).

20 Roma – ODD Customization, presentation and documentation: <https://roma.tei-c.org/> (last accessed: 19 November 2023).

are relevant for the whole corpus and as they account for the way it has been worked on as a whole, the technical presentation and access to the full documentation are located in the “encodingDesc” of the corpus file (ACP03\_annuaires.xml). This full documentation showing the whole structure and listing the elements and attributes can be found in spreadsheet files (in CSV format: 6 files; in Excel format: 1 file with 6 tabs).<sup>21</sup> It is intended to be both a guide and a dictionary for researchers who would like to continue the encoding.

## (4) RESULTS AND DISCUSSION

### (4.1) DATA ARCHITECTURE AND STRUCTURE

As introduced in the second part of this paper, our work resulted in three separate datasets linked by the indication of relations between them and by the common prefix “ACP03”<sup>22</sup> in the file names. The main one gathers the XML-TEI files, along with the RNG schema files associated with them (ACP03\_teiAnnuaire.rng; ACP03\_teiAnnuaire\_1968.rng); the XSLT files which enabled the production of the two other datasets; and the documentation files (CSV and Excel).

The XML-TEI file ACP03\_annuaires.xml contains a main “teiCorpus” element. In the “sourceDesc”, it lists all the yearbooks which have been encoded and the documents that have been withdrawn from the corpus for this period of study (“bibl[@status=‘withdrawn’]”). With “xi:include” tags, this file includes the various parts of ACP03\_index\_personnes.xml (in the “profileDesc”) and the encodings of the yearbooks. All the XSLT stylesheets can be applied directly at this level.

Each XML-TEI file is named with the pattern “ACP03\_oas[YYYY].xml” (i.e. “ACP03\_oas1883.xml”). The main information about the source and the main steps of its encoding are indicated in the metadata section (“teiHeader”). The “body” section embeds all the text worked on, whereas the “front” and “back” sections are filled with introductory parts and annexes roughly transcribed in XML comments. Though referenced by the “pb” elements of these files, the numerous IMG files of the digitalisations have not been put in open access in order to encourage interested researchers to contact the Archival City team and the Archives de Paris, and to consult both the generated PDFs and the visualization tool (see below).

The pattern of this corpus, the elements that were used and the uses that were made of them are documented: in CSV format, in the files “ACP03\_doc1\_archiGe.csv” (for the general architecture of the corpus), “ACP03\_doc2\_annuaireXml.csv” (for the file “ACP03\_annuaires.xml”), “ACP03\_doc3\_indexPersonnes.csv” (for the file “ACP03\_index\_personnes.xml”), “ACP03\_doc4\_oasAAAA\_teiHeaderStruct.csv” (for the teiHeader and the general structure of the files of each yearbook), “ACP03\_doc5\_oasAAA\_body.csv” (for the content of the “body” elements of the files of each yearbook), and “ACP03\_doc6\_distinctions.csv” (for the transcription conventions used for the representations of honorary distinctions); in XSLX format, in the “ACP03\_doc.xlsx” file, with a corresponding tab for each of the CSV files mentioned.

The second dataset contains PDF exports crossing the XML-TEI data and the images (name pattern: ACP03\_oas[YYYY]\_img.pdf). It always starts with the metadata of the encoding, then it continues with the clickable table of contents of the PDF document, with the (also clickable) table of contents of the source in the form of the hierarchy encoded from it (Figure 4). It refers to the digitisations of the yearbook which are displayed after it. This same table of contents and hierarchy is accessible in the PDF navigation pane (Figure 5).

The third dataset gathers the Markdown and CSV files generated with XSLT stylesheets. The Markdown files (name pattern: ACP03\_oas[YYYY].md) display only the hierarchies of the yearbooks (values of the “orgName” elements put after as many “#” as the number of “org” elements they are embedded in). They can be imported in mind map software.

---

<sup>21</sup> This documentation does not account for the small specificities of the schema of the 1968 file, which are quite easy to understand.

<sup>22</sup> “ACP03” standing for “Archival City Greater Paris”, 3<sup>rd</sup> group/set of data.

1955 - Table des matières de l'annuaire / Hiérarchie - Nomenclature sommaire des services de la préfecture de la Seine (octobre 1955)

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| III — Bureau des loyers, des échanges et des primes de déménagement             | 38 |
| III — Bureau de l'habitat et du contrôle des démolitions d'immeubles            | 38 |
| IV — Bureau du contrôle des transformations de locaux et des cotes d'immeubles  | 38 |
| V — Bureau de la taxe de compensation sur les locaux insuffisamment occupés     | 38 |
| VI — Bureau du contentieux des réquisitions                                     | 38 |
| VII — Bureau des attributions et des notifications                              | 38 |
| DIRECTION DU PERSONNEL                                                          | 39 |
| Secrétariat                                                                     | 39 |
| SOUS-DIRECTION DES AFFAIRES GÉNÉRALES DE PERSONNEL                              | 39 |
| I — Bureau du statut                                                            | 39 |
| Service d'ordre                                                                 | 39 |
| Secrétariat du Conseil de discipline                                            | 39 |
| III — Bureau des traitements et indemnités                                      | 39 |
| III — Bureau des pensions et secours                                            | 39 |
| IV — Ecole d'administration de la Ville de Paris                                | 39 |
| SOUS-DIRECTION DE L'ADMINISTRATION DES PERSONNELS                               | 40 |
| I — Bureau des personnels administratifs et de service                          | 40 |
| INSPECTION DU PERSONNEL DE SERVICE                                              | 40 |
| II — Bureau des personnels technique et ouvrier                                 | 40 |
| III — Bureau des personnels spéciaux                                            | 40 |
| IV — Bureau de la comptabilité générale du personnel et centre de mécanographie | 40 |
| a) SECTION DE LA COMPTABILITÉ DU PERSONNEL                                      | 40 |
| b) SECTION DES PRESTATIONS FAMILIALES ET SOCIALES                               | 40 |
| c) CENTRE DE MÉCANOGRAPHIE                                                      | 40 |
| V — Service médical                                                             | 40 |
| DIRECTION DES PONTS ET CHAUSSEES ET DES TRANSPORTS                              | 41 |
| SECRETARIAT                                                                     | 41 |
| BUREAU                                                                          | 41 |
| I — PONTS ET CHAUSSEES                                                          | 41 |
| A — Service ordinaire et vicinal                                                | 41 |
| 1er arrondissement (Nord)                                                       | 41 |
| SUBDIVISIONS                                                                    | 41 |
| Courbevoie                                                                      | 41 |
| Asnières                                                                        | 41 |

0 1 2 3 4 5 6 16

Figure 4 Edition of 1955, PDF export, table of contents.

1955 - Nomenclature sommaire des services de la préfecture de la Seine (octobre 1955) - Pages numérisées et transcrites

— 41 —

Direction des ponts et chaussées et des transports

**DIRECTION DES PONTS ET CHAUSSEES ET DES TRANSPORTS**

7, rue Beaubourg (r. T. 44-20 à 23, 71-73 à 74; Ann. 36-14 et 15  
(sauf les services ayant une ligne spéciale).)

M. LAFFRÈRE (L.), ingénieur en chef des ponts et chaussées, directeur.

M. DURUZI (Y.), ingénieur en chef des ponts et chaussées.

M. BERTEAU (R.), ingénieur en chef des ponts et chaussées.

M. MOYER (F.), ingénieur en chef des ponts et chaussées.

**SECRETARIAT**

M. LE VIVANT, ingénieur des travaux publics de l'Etat.

**BUREAU**

M. LUCIANI, ingénieur des travaux publics de l'Etat.

M. FELLMEI, ingénieur des travaux publics de l'Etat.

M. FLECKE, ingénieur des travaux publics de l'Etat.

**I — PONTS ET CHAUSSEES**

**A. — Service ordinaire et vicinal**  
(Quatre arrondissements)

1<sup>er</sup> arrondissement (Nord)  
 7, rue Beaubourg.

M. FURAND-DURHEF, ingénieur des ponts et chaussées.

M. BROUCCHE, ingénieur T. P. E., chef de bureau.

**SUBDIVISIONS**

Courbevoie, 20, boulevard de la Minette-Marchand :  
 M. BELLETIER, ingénieur T. P. E.  
 Pl. 12-57

Clichy, 14, rue Vanuxem :  
 M. MELLERON, ingénieur T. P. E.  
 Pl. 13-29

Neuilly-sur-Seine, 179, avenue de Préville-Whites :  
 M. COSTANT, ingénieur T. P. E.  
 Pl. 33-55

Le signe @ indique le service de base (arrondissement).  
 Le signe S indique les services d'appel téléphoniques relatifs aux arrondissements Ann. 53-25.  
 Le signe T indique les services d'appel téléphoniques relatifs à d'autres arrondissements.

**Signets**

- MATERIEL ET DU CONTR...
- DIRECTION DU PERSONNEL (p. 39)
- ▼ DIRECTION DES PONTS ET CHAUSSEES ET DES TRANSPORTS (p. 41)
  - SECRETARIAT (p. 41)
  - BUREAU (p. 41)
  - ▼ I — PONTS ET CHAUSSEES (p. 41)
    - ▼ A. — Service ordinaire et vicinal (p. 41)
      - ▼ 1er arrondissement (Nord) (p. 41)
        - ▼ SUBDIVISIONS (p. ...
          - Courbevoie (p. 41)
          - Asnières (p. 41)
          - Clichy (p. 41)
          - Saint-Denis (p. 41)
          - Neuilly-sur-Seine (p. 41)
          - Puteaux (p. 41)
          - Aménagement (p. ...
        - ▼ 2e arrondissement (Est) (p. 41)
          - ▼ SUBDIVISIONS (p. ...
            - Aubervilliers (p. 41)
            - Pantin (p. 41)
            - Neuilly-le-Sec (p. 41)

Figure 5 Edition of 1955, p.41, PDF export.

The CSV files with the name pattern “ACP03\_oas[YYYY]\_lieux.csv” are exports of the references to office addresses, with one line per “placeName” element. The file “ACP03\_annuaires\_lieux.csv” gathers this information for all the yearbooks. The type of information all these files contain is detailed in the following table (Table 1).

The CSV files with the pattern “ACP03\_oas[YYYY]\_personnes.csv” are exports of the references to individuals, with one line per “persName” element. The file “ACP03\_annuaires\_personnes.csv” gathers this information for all the yearbooks. The file “ACP03\_doc6\_distinctions.csv” has been copied in this dataset to enable understanding the “Distinctions honorifiques” column. The type of information all these files contain is detailed in the following table (Table 2).

| COLUMN HEADINGS (IN FRENCH)         | COLUMN CONTENT                                                                  |
|-------------------------------------|---------------------------------------------------------------------------------|
| Année                               | year of the yearbook concerned                                                  |
| Mention complète                    | complete passage encoded in the “placeName”                                     |
| Chemin dans l’annuaire              | XML-TEI hierarchy in which this reference is embedded, from top to lowest level |
| Page de l’annuaire : numéro         | page number written in the source                                               |
| Page de l’annuaire : nom de fichier | file name of the digitalization of this page                                    |

**Table 1** Detail about the content of the CSV files “ACP03\_oas[YYYY]\_lieux.csv” (one line per “placeName”).

| COLUMN HEADINGS (IN FRENCH)         | COLUMN CONTENT                                                                                               |
|-------------------------------------|--------------------------------------------------------------------------------------------------------------|
| Année                               | year of the yearbook concerned                                                                               |
| Identifiant                         | identifier of this person if listed in the index file                                                        |
| Nom (particule)                     | last name of this person, with its possible name link, if listed in the index file                           |
| Prénoms                             | first names of this person, if listed in the index file                                                      |
| Sexe                                | sex of this person, if listed in the index file                                                              |
| Mention complète                    | complete passage encoded in the “persName”                                                                   |
| Distinctions honorifiques           | content of the embedded “roleName[@type=’distinction’]” element (for honors), if present                     |
| Titres                              | content of the embedded “roleName[@type=’titre’]” element, if present (for professional titles)              |
| Chemin dans l’annuaire              | XML-TEI hierarchy in which this reference is embedded, from top to lowest level                              |
| Position hiérarchique               | content of the embedded “affiliation[@type=’position-hierarchique’]”, if present (for hierarchical position) |
| Attributions directes               | content of the embedded “State” element, if present (for details on the assignments of this person)          |
| Lieu                                | content of the embedded “placeName” element, if present                                                      |
| Informations bureau                 | content of the embedded “district” element, if present (for office, telegraph and phone numbers)             |
| Page de l’annuaire : numéro         | page number written in the source                                                                            |
| Page de l’annuaire : nom de fichier | file name of the digitalization of this page                                                                 |

**Table 2** Detail about the content of the CSV files “ACP03\_oas[YYYY]\_personnes.csv” (one line per “persName”).

## (4.2) DATA LIMITATIONS

Although they allow finding information easily and quite synthetically, these datasets must be consulted with a discerning eye and with a willingness to consider them as a basis for further analysis, even if it means rethinking a part of our encoding.

As the layout and typographies of the titles in the yearbooks may give little information or be misleading about how the departments are organized between them, interpretations had to be made in the elaboration of the hierarchies of “org” elements. Likewise, the certainty about what titles are department names or not is moderate. Because of this, the hierarchies of the XML-TEI and export files may contain errors due to the necessity of gathering more knowledge about these administrations through other sources.

Moreover, parts of some yearbooks do not appear in XML, with a transcription not corrected, and typing errors which may remain elsewhere. This may lead to irrelevant text research in cases when a word is written in the source but not encoded or transcribed properly. Thus, keeping in mind that not all the yearbooks have been worked on entirely is crucial when looking for a specific word, place or individual. In a general manner, the users must keep in mind that these encodings have been produced manually, with all the precision and defects this may involve.

Regarding the CSV exports, especially those related to persons, it is important to acknowledge that some contents extracted in each line are only those embedded in the references

concerned (see above). However, the complete information about a person, for example, cannot be put in the “persName” element integrally. For example, this occurs when there is a list of physicians (or others), with the word “Médecins” at the its head. The link between this function and each name would require complex encoding that we did not have the time to do, and a subsequent complex code in XSLT to manage all these different types of encodings.

In a general manner, we did not have the means or the time to encode the yearbooks to their full potential. The constraint of research functionalities could be overcome by the continuation of our work in multiple ways.

## (5) IMPLICATIONS/APPLICATIONS

We hope that our work will be useful to the scientific community for many purposes. The following paragraphs highlight possible reuses and implications, and present the visualisation tool which will be available at the beginning of 2024.

### (5.1) FOR FURTHER ENCODINGS AND INVESTIGATIONS

Historians are most welcome to use and reuse these datasets, for consultation, compilation and edition. In the XML-TEI documents, consultation can be enhanced by XPath queries on various subjects. For example, such a query returns the editions of the yearbooks mentioning the rationings related to World War II (1942, 1945, 1948 and 1949): `//TEI//orgName[contains(.,“rationnement”)][1]/ancestor::TEI//titleStmt/title/date`.

Many promising advances can be made in the encoding of these yearbooks by other teams and researchers.

The first one would be to link more closely the yearbooks between them by putting specific markers on each department, to follow their evolution through time and their genealogy (mergers, division, etc.). This has already been done for a few departments such as the public road departments (“voie publique”/“voirie”) with the addition of a “@role” attribute to the “org” elements concerned. If there is no need for a more complex technical approach, this method could be applied to all the other types of departments while the necessary historical analysis is performed.

The objective of creating a historical thesaurus of functions and assignments could be completed, with the related aim of better tracing and understanding the evolution of the administration. Thus, this thesaurus should be used to tag the departments in the yearbooks. These are ambitious tasks, but they would provide the opportunity to investigate more broadly and more closely a part or the whole of the city’s administration.

Obvious continuities would be to widen our index of persons and to gather more external information in it, such as links to prosopography databases. Furthermore, one of the investigative paths which had to be left aside because of lack of time was to reference the office addresses geo-reference them, and tag them for this very purpose (going farther than a neutral “placeName” element). The ambition of these possible tasks stems as much from the extent of the yearbooks as from the necessity to refer to places and streets that no longer exist and to capitalize on the current reference systems.<sup>23</sup>

Finally, a very interesting continuity would be to link the departments listed in the yearbooks with archival descriptions and authority records (yet to be elaborated). This would give the possibility to access archives (through the search aids) and synthetic historical notes from the same yearbooks, and vice versa.

### (5.2) THE VISUALISATION WEBSITE

As a showcase of our work and of the potentialities of these yearbooks, the visualisation tool available at the beginning of 2024 will compile the data of the XML-TEI files, give access to the digitalisations and let the user navigate through the chronology and the hierarchies of the documents.

---

<sup>23</sup> I.e. the reference systems of the streets of Paris put on open access by the Archives Nationales (France): [https://github.com/ArchivesNationalesFR/Referentiels/blob/main/lieux/csv/Fran\\_Paris\\_Voies.csv](https://github.com/ArchivesNationalesFR/Referentiels/blob/main/lieux/csv/Fran_Paris_Voies.csv).

This website will display two main types of access: a general chronology of the yearbooks and the first levels of the hierarchies they show; a catalogue of the yearbooks with search functions for transcribed words. These two entrances will lead to the tables of contents of each yearbook and to the display of each page, with the digitalization and the transcription. Zoom and download functions, page by page, will allow the user to enter the source. Obviously, the search functions will be as constrained as the CSV exports are.

The aim of this website will be to facilitate access to the yearbooks as much as to draw the attention of the scientific community to these crucial sources of administrative history.

## ACKNOWLEDGEMENTS

Vincent Lemire: director of the Archival City project.

Loïc Vadelorge: director of the Greater Paris research field of the Archival City project and the administrative yearbooks analysis.

Other researchers involved: Emmanuel Bellanger, Cédric Fériel, Paul Lecat, Paul Lesieur,

Archives de Paris: Guillaume Nahon, Béatrice Hérold, Vincent Tuchais, Gaël Donneger, Nicolas Courtin.

Bibliothèque de l'Hôtel de Ville: Renaud Fuchs.

Ecole des Ponts ParisTech: Charles Riondet, Anne Lacourt.

Cassandra Maubert: principal curator of the administrative yearbooks encoding.

Other curators of the administrative yearbooks encoding: Sonya Bensaadi, Fanny Brière, Nathan Dacier-Falque, Valentin Devillepoix, Andréa Dorin, Jeanne Fras, Anahi Haedo, Léna Humbert.

## FUNDING INFORMATION

Research funded by the Archival City project of the I-Site FUTURE, Université Gustave Eiffel (France).

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

Carole Lamoureux: conceptualization, data curation, methodology, supervision, visualization, writing – original draft

Elsa Camus: conceptualization, data curation, methodology, writing – original draft

## AUTHOR AFFILIATIONS

**Carole Lamoureux**  [orcid.org/0000-0002-6246-0227](https://orcid.org/0000-0002-6246-0227)

Equipe d'accueil Analyse comparée des pouvoirs (ACP), Université Gustave Eiffel, Champs-sur-Marne, France

**Elsa Camus**

Library, Assemblée nationale, Paris, France

## REFERENCES

**Archival City.** (2019–2023). Metropolis Archives. *Archival City*. Retrieved from <https://archivalcity.hypotheses.org/the-project/grand-paris/metropolis-archives> (last accessed: 19 November 2023).

**Chagué, A.** (2022). eScriptorium : une application libre pour la transcription automatique des manuscrits. *Arabesques*, 107, 25. DOI: <https://doi.org/10.35562/arabesques.3100>

**Fekete, J.-D., & Ogilvie, D.** (2008). Le projet et la plateforme “millefeuille”. Recherches et outils informatiques pour de nouveaux usages des almanachs administratifs. In B. Delmas, D. Margairaz

& D. Ogilvie (Eds.), *Bibliothèque de l'Ecole des chartes*, vol. 166. DOI: <https://doi.org/10.3406/bec.2008.464662>

**Lesieur, P.** (2023). A Database for studying the public roads department of the Préfecture de la Seine (19th–20th centuries). *Archival City*. Retrieved from <https://archivalcity.hypotheses.org/4822> (last accessed: 19 November 2023).

**Vadelorge, L.** (2024). Penser le Grand Paris par les annuaires administratifs : l'exemple de la planification urbaine. *Histoire Urbaine*, (forthcoming).

Lamoureux and Camus  
*Journal of Open  
Humanities Data*  
DOI: 10.5334/johd.186

15

**TO CITE THIS ARTICLE:**

Lamoureux, C., & Camus, E. (2024). Moving Forward in Administrative History: Encoding the Département de La Seine and Paris Yearbooks (1883–1970). *Journal of Open Humanities Data*, 10: 18, pp. 1–15. DOI: <https://doi.org/10.5334/johd.186>

**Submitted:** 20 November 2023

**Accepted:** 20 December 2023

**Published:** 13 February 2024

**COPYRIGHT:**

© 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Journal of Open Humanities Data* is a peer-reviewed open access journal published by Ubiquity Press.