



**HAL**  
open science

# LectAuRep : Données d'archives en français des XIXe et XXe siècles

Hugo Scheithauer

## ► To cite this version:

Hugo Scheithauer. LectAuRep : Données d'archives en français des XIXe et XXe siècles. Transkribus / eScriptorium : Transcrire, annoter et éditer numériquement des documents d'archives, Bibliothèque nationale de France, May 2022, Paris, France. hal-03666884

**HAL Id: hal-03666884**

**<https://inria.hal.science/hal-03666884v1>**

Submitted on 12 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

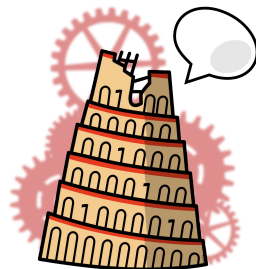
# LectAuRep

Données d'archives en français des XIXe et XXe siècles

Transkribus / eScriptorium : Transcrire, annoter et éditer numériquement des documents  
d'archives (09/05/22)

Hugo Scheithauer (Inria)

*Inria*



# LectAuRep

Inria - Archives nationales (Département du Minutier Central)

- Faciliter et massifier la consultation des répertoires de notaires conservés au DMC grâce aux technologies de reconnaissance d'écriture manuscrite.
- Offrir aux usagers une lecture enrichie des répertoires et permettre la fouille de données dans le corpus.

Voir Alix Chagué, Rostaing Aurélia. LECTAUREP: Paris Notary Record Books Automated Reading. Fantastic Futures 2021 / Futures Fantastiques 2021, AI4LAM; BnF; Université Paris Saclay, Dec 2021, Paris, France.

N <sup>os</sup> DU RÉPERTOIRE	DATES DES ACTES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES  INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.		
		EN BREVETS	EN MINUTES		DATES	DROITS	
<b>An 1911 , mois d' Août</b>							
733	18	Mainlevée	(Subst. M <sup>e</sup> Moreau)	Sebourg (par Maria Rigaudat N <sup>e</sup> d'Amable) d'Asnières, routes, des bois arrières les mains de l'ancien	4	"	"
734	18	(Subst. M <sup>e</sup> Moreau)	Quittance	Sebourg à la d. dame par Demoucheux, d'Asnières, de 11106 <sup>f</sup> 73	"	"	"
735	18	( - 2 <sup>e</sup> - )	Dépôt de décharge	Dubuffet demandeur par Paul Lucien, sa femme Bayonne, à M. M. Lepur et Brochelain	"	"	"
736	18	(Subst. M <sup>e</sup> L. Baudrier)	Dépôt de testament	Devillers de Marie Joseph f <sup>e</sup> divorcée, à Paris, rue Viollet la Rue 5	"	"	"
737	19	Procuration		Guignan (par Hermann) et Gabrielle Françoise Pauline Faure, ép <sup>ux</sup> à Paris. Bd			
738	19		Obligation	Magenta, 118, se acquies par licitation un immeuble à Bordeaux, rue Lagrange, 99	19	3	75
739	19		- 2 <sup>e</sup> -	Sterckeman (par Jules Charles Bertin Lucien) et Lucie Aimée Secusble, ép <sup>ux</sup> à Paris, Bd Lames, 31 <sup>er</sup> et Adolphe Weil, rue Beaurepaire 27, de 22000 <sup>f</sup>	22	350	"
740	19	(Subst. M <sup>e</sup> Moreau)	Décharge de legs	- 2 <sup>e</sup> - (Charles mêmes) à Gabrielle Zoé Thérèse Henriette Brunelle, V <sup>e</sup> de son Marcellin Albert Petit, à Paris, faubourg St-Denis, 210, de 30000 <sup>f</sup>	22	375	"
				Vareille (par Marie) f <sup>e</sup> divorcée, 33 Bd de la Seine, à la Noue Louis Jean Alfred Renaud	"	"	"

Extrait d'une page d'un répertoire de notaire.

Transcrire les répertoires de notaires

N <sup>os</sup> DU RÉPERTOIRE	DATES DES ACTES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES  INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.		
		EN BREVETS	EN MINUTES		DATES	DROITS	
<b>An 1911 , mois d' Août</b>							
733	18	Mainlevée	(Subst. M <sup>e</sup> Moreau)	Lebon (par Maria Papaudat l <sup>e</sup> d'Amable) d'Asnières, et autres, de seize arres sur les murs de l'ouveau	4	"	"
734	18	(Subst. M <sup>e</sup> Moreau)	Quittance	Lebourg (à la d. dame) par Beaumont & Co <sup>e</sup> , d'Asnières, de 4406 <sup>fr</sup> 73	"	"	"
735	18	(2 <sup>e</sup> )	Déclat de décharge	Dubuffet demandeur par Paul Lucien, s <sup>o</sup> f. d <sup>e</sup> Bayonne, à M. M. Lefeur et Bachelain	"	"	"
736	18	(Subst. M <sup>e</sup> L. Baudrier)	Déclat de testament	Devillers (de Felicie Joseph) f. divorcée, à Paris, rue Vollet la Duc 5	"	"	"
737	19	Procuratlon		Guignan (par Hermann) et Gabrielle Françoise Pauline Faux, ép <sup>x</sup> à Paris. Bd <sup>e</sup> Magenta, 148, pr acquisition par licitation un immeuble à Bordeaux, rue Lagrange, 99	19	3	75
738	19		Obligation	Sterckeman (par Jules Charles Bertin Lucien) et Lucie Amélie Levasseur, ép <sup>x</sup> à Paris, Bd <sup>e</sup> James, 31 <sup>bis</sup> à Adolphe Vêl, rue Beaurepaire 27, de 28000 <sup>fr</sup> ..	22	350	"
739	19		- 2 <sup>e</sup> -	- 2 <sup>e</sup> - (par les mêmes) à Gabrielle Zoé Flora Henriette Brunelle, V <sup>e</sup> de Lion	22	375	"
				Marcellin Albert Petit, à Paris, faubourg St <sup>e</sup> Denis, 210, de 30000 <sup>fr</sup> ..			
740	19	(Subst. M <sup>e</sup> Moreau)	Décharge de legs	Vareille (par Marie) f. divorcée, 32 Bd <sup>e</sup> de la Chapelle, à la H <sup>o</sup> u <sup>e</sup> Louis Jean Alfred Renault	"	"	"
741	19		Inventaire	Garnier (après décès de M. l'abbé Henri Joseph f. B <sup>e</sup> pr vic <sup>e</sup> de l'Église St <sup>e</sup> Genevieve des Grandes Carrières, 2 <sup>e</sup> à Paris, rue Championnet 174, décès au hôpital Richelieu 20 juillet 1911	21	7	50



S ET DOMICILES DES PARTIES

SITUATIONS ET PRIX DES BIENS

RELATION  
de  
l'Enregistrement.

Nos	DATES	Noms	NATURE ET ESPÈCE	SITUATIONS ET PRIX DES BIENS	RELATION de l'Enregistrement.	
					DATES	DROITS
292	15	P.V. o.u.v. Col.		An 1939, mois de Mars Bajul au décès de Désiré Alexandre Marais à Romaniville get de uedes Oserais arrivé le 10 <sup>8</sup> 1938	22	-
293	15		Inventaire	Machinau décès de Dieudonné Joseph Adolphe à Paris 6 rue des Batignolles y arrivé le 22. 12-1938 ep. de Henriette Bugéine Moutigny	22	35 -
294			Vente	Perrin et Anderson à Jules Menny à Dravey 11 rue Darcy de 197 <sup>4</sup> 02 terrains etoile Dravey n° 103 du plan S <sup>m</sup> A.P. 40110 prix 9756,65	23	1747 04
295		17 Certif. sign.		Baer couit Berthe Isabelle Henriette Bourdin ep. de Richard Philippe Auché à Paris 10 <sup>8</sup> de Courcelles	23	35 -
296	17	Promotion		Coté J. Raymond à Paris 6 rue S'Amsterdam en l.p. l. venche	23	35 -

- Deux modèles génériques avec un taux d'erreur par caractère inférieur à 10%.
- Des modèles spécialisés qui résultent de l'affinage des modèles génériques sur des lots de données plus petits et plus uniformes, atteignant un taux d'erreur par caractère égal voire inférieur à 5%.



Conserver la structure logique des répertoires

Nos	DATES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES	RELATION DE	
		BREVETS	EN MINUTES		Enregistrement	DROITS
				INDICATIONS, SITUATIONS ET PRIX DES BIENS	DATES	
505	21		Donation	An 1901, mois de Mai		
506	21		Dr	De (par Louis Chédoré) à Paris, rue Berniquin 60, à Louise Clément, sa femme	"	"
507	21		Notoriété	De (par M <sup>me</sup> ) son à son mari	"	"
				Mariés (après décès de Marie Louise Berthe), décédée à Paris (Eaubou) du 22		
508	21		Procuration	Mars 1900	5	3,75
				De Kirschen (par Louis Eugénie) à Neuilly, 1 <sup>er</sup> étage, rue de Labouville,		
509	21		Procuration	30 <sup>ème</sup> 1 <sup>er</sup> Not. Mouchet Vaccovère	1	3,75
				Guillaumin (par Louis Eugénie) à Neuilly, 1 <sup>er</sup> étage, rue de Labouville,		
510			Certificat	Mois de Juin 1901	7	3,75
			1. propriété	Legenre (après décès de 1500 <sup>fr</sup> au nom de Rose Etol, 1 <sup>er</sup> Not. Martial Paul Eugénie)		
511	1		Notoriété	chicé	1	Judis
512	1		Procuration	Dr (après le décès du dit) à Levallois Perret, rue du Bois 65, du 11 <sup>ème</sup> 1900	1	3,75
				Oubateq (par Justine Adèle Robinson, 1 <sup>er</sup> Not. Dubois) à Paris, rue de la		
				20 <sup>ème</sup> 1 <sup>er</sup> Not. comarthe maréchal	4	3,75

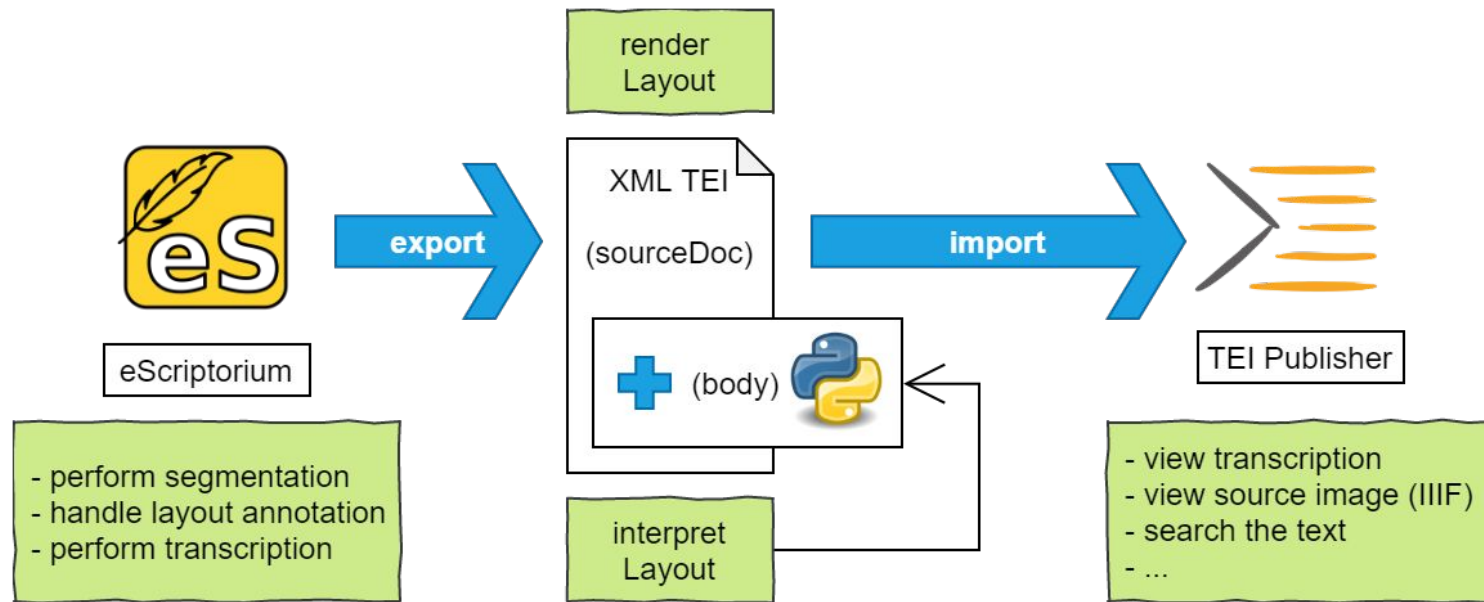
Détection des lignes de texte dans un répertoire de notaire

56 Procuration  
57 Audtorisation  
58 Certificat de ppte  
59 Procuration  
60 Mainlevée  
61 de signification  
62 Mainlevée  
63 Mainlevée  
64 Mainlevée  
65 Mainlevée  
66 Obligation  
67 Dépôt  
68 Transport  
69 Vente  
70 Constitution  
71 deSequestre  
72 Décharge  
73 Procuration  
74 Substitution  
75 Quittance  
76 Vente  
77 An 1914, mois d. Fevrier  
78 à Paris 34 rue Sedaine  
79 Lamour (par Jn B^e) employé à la Caserne de la G^de Républicaine  
80 à Paris rue St Etienne du Mont en blanc pour recueillir SS^on  
81 Boettcher (par Henry Emile) Ing^t conseil à Paris 39 Bd St Martin  
82 à Léontine Jeanne Blanchot safe dt avec lui pour gérer Comce  
83 Boileau-Desombres (parlaSté) à Paris 73 rue Lafayette d'inxn c/  
84 Sophie Marie Louise Dessoliers épse de Edouard Lavelaine de Maubeuge  
85 Gront (par Jean Eugène André) direstt d'usme à Bahigny route des  
86 Petits Ponts à André Valern Fortin ind^el à Paris 34 rue Sedaine

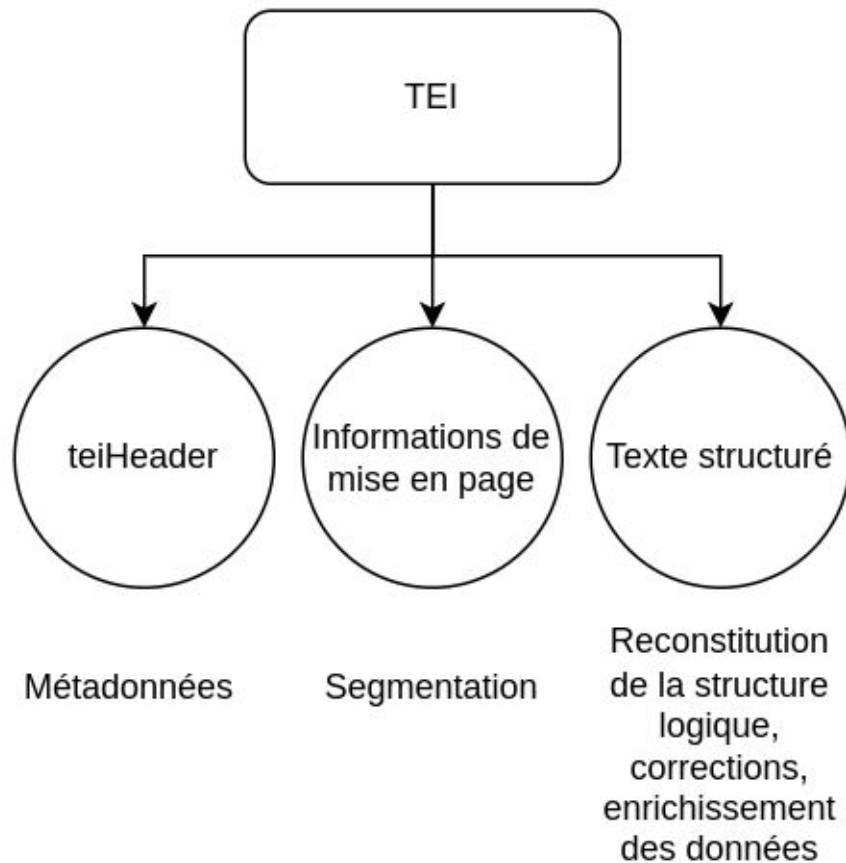
Extrait d'une transcription à plat,  
exportée en texte brut.

N <sup>os</sup> DU RÉPERTOIRE	DATES	NATURE ET ESPÈCE DES ACTES		INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.	
					DATES	DROITS
1358	8	Certif <sup>te</sup> de pp <sup>te</sup>		An 1927, mois d'Août Baladilhe (concern <sup>t</sup> 5 extraits d'inscrip <sup>on</sup> au total de 1419 <sup>x</sup> de rente f <sup>se</sup> 3% au même nom que ci-dessus	11	22 50
1359	8	d <sup>o</sup>		Baladilhe (concern <sup>t</sup> divers extraits d'inscrip <sup>on</sup> de rente f <sup>se</sup> au nom de Lefebvre - Desvallières Jeanne Elisabeth Georgina, f <sup>e</sup> de Emile / <i>subnommé</i>	11	22 50
1360	8	d <sup>o</sup>		Baladilhe (concern <sup>t</sup> divers extraits d'inscrip <sup>on</sup> de rente f <sup>se</sup> au même nom que ci-d <sup>essus</sup>	11	22 50
1361	8	(son M <sup>e</sup> Delarue)	Mainlevée	Blonde et Boigaud (par la S <sup>te</sup> / de Paris 14 rue de Nâcon - d'inscrip <sup>on</sup> cf Raphaël Boigaud et H <sup>e</sup> ép <sup>te</sup> de Paris 24 f <sup>ce</sup> du Marché St Honoré		
1362	8	Procuration	(son M <sup>e</sup> Delarue)	Wathoz (par Fernand / de Paris 113 rue du Chemin Vert en blanc - pour vendre		
1363	8	Mainlevée	d <sup>o</sup>	Tricotel (par Georges / de Paris 8 bd Poissonnière - d'inscrip <sup>on</sup> cf André Lécraud et H <sup>e</sup> ép <sup>te</sup> de Paris 207 rue de Crimée		

Détection des différentes régions de texte dans un répertoire de notaire



La chaîne de traitement LEPIDEMO



N <sup>os</sup> DU RÉPERTOIRE	DATES	NATURE ET ESPÈCE DES ACTES		INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.	
		ARTICLES	2 DÉCIMES EN SUS 1927		DATES	DROITS
1358	8	Certificat de pp <sup>te</sup>		An 1927, mois d'Avril Baladilhe (concernant 5 extraits d'inscriptions au total de 1419 <sup>x</sup> de rente 3 <sup>o</sup> / <sub>100</sub> au même nom que ci-dessus	11	22 50
1359	8	d <sup>o</sup>		Baladilhe (concernant divers extraits d'inscriptions de rente 3 <sup>o</sup> / <sub>100</sub> au nom de Sébastien Desvallières Jeanne Elisabeth Georgina, f <sup>e</sup> de Emilie / subnomme	11	22 50
1360	8	d <sup>o</sup>		Baladilhe (concernant divers extraits d'inscriptions de rente 3 <sup>o</sup> / <sub>100</sub> au même nom que ci-dessus	11	22 50
1361	8	(son N <sup>o</sup> Delarue)	Mainlevée	Blonde et Boigaud (par la 3 <sup>o</sup> / <sub>100</sub> de Paris 14 rue de Mécon d'inscriptions et 4 <sup>o</sup> / <sub>100</sub> de Paris 24 p <sup>o</sup> du Marché St Honoré		
1362	8	Procuration	(son N <sup>o</sup> Delarue)	Wathiez (par Fernand / de Paris 113 rue du Chemin Vert en blanc pour vendre		
1363	8	Mainlevée	d <sup>o</sup>	Tricotel (par Georges / de Paris 8 bd Poissonnière d'inscriptions et André Lévraud et 4 <sup>o</sup> / <sub>100</sub> de Paris 207 rue de Crimée		
1364	9	Certificat de vic	d <sup>o</sup>	Bourcier (concernant Louis / de Paris 43 imp <sup>o</sup> du Poussieu		

Reconstituer des unités logiques pour signaler l'information avec des traitements par règle ou avec de la reconnaissance d'entités nommées.

Numéros du répertoire	Dates des actes	Actes en brevets	Actes en minutes	Noms, prénoms et domiciles des parties ; indication, situations et prix des biens	Date de l'enregistrement	Droits de l'enregistrement	Autres
292	15	P. V. ouv. Cof.		Bajul au décès de Désiré Alexandre Marius à Romainville 7 et 9 rue des Oseraies arrivé le 10 8 <sup>bre</sup> 1938	22	-	
293	15		Inventaire	Jadin au décès de Dieudonné Joseph Adolphe à Paris 6 rue des Batigolles y arrivé le 22-12-1938 ep <sup>^</sup> x de Henriette Eugénie Montigny	22	35 -	
294		16	Vente	Pierrat et Anderson à Jules Menny à Drancy 11 rue Daisy de 297 <sup>^</sup> f 02 terrain étoile Drancy n°203 du plan XXX A. P n°110 prix 9756,65	23	1747 04	
295		17 Certif Sign.		Baer conct Berthe Isabelle Henriette Bourdin ep <sup>^</sup> se de Richard Philippe André à Paris 20 B <sup>^</sup> d de Courcelles	23	35 -	



# Conclusion

- La segmentation est essentielle pour bien délimiter des unités de texte, mais également pour comprendre ces unités une fois transcrites au sein de la mise en page.
- La transcription permet d'obtenir la donnée, et plus un modèle sera robuste, plus la donnée sera de qualité.

# Déployer eScriptorium à Inria

- 2019 : machine virtuelle, sans carte graphique.
- 2020 : serveur traces6, 2 cartes graphiques, plus de stockage.
- 2021-2022 : CREMMA, 2 cartes graphiques supplémentaires, plus de stockage, plus de mémoire vive.

**Merci pour votre attention !**