DataCatalogue : enjeux et réalisations

Laurent Romary, Hugo Scheithauer

DataCatalogue

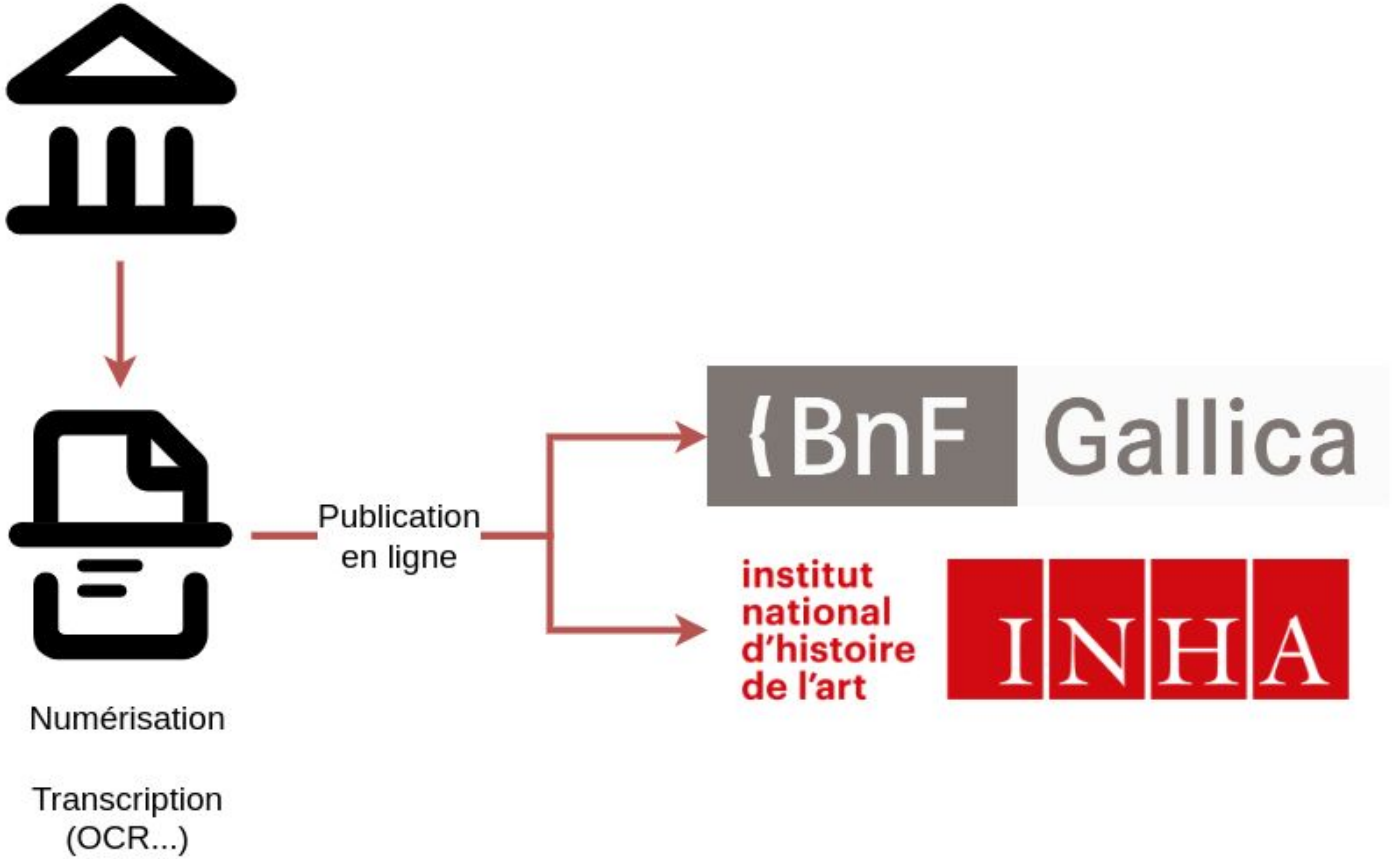
Enjeux et réalisations

Un outil numérique pour interroger les catalogues de vente : le projet DataCatalogue - 21/10/22

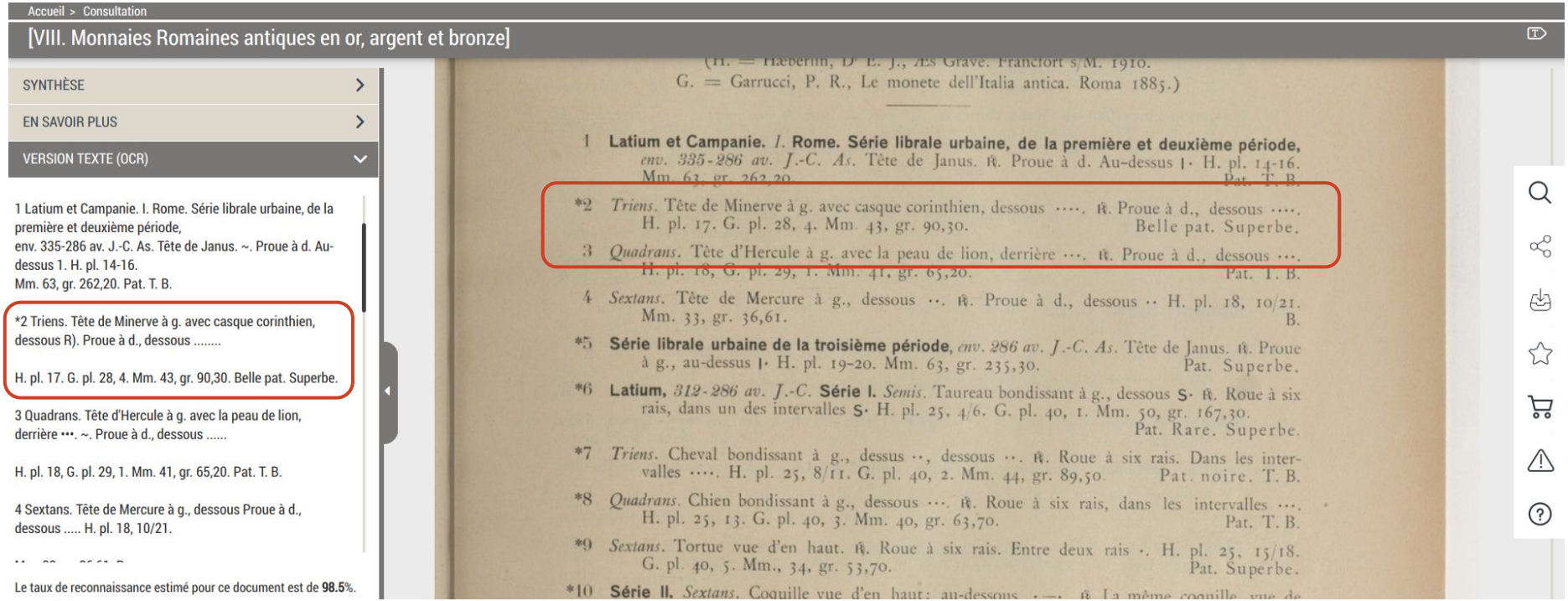
Laurent Romary (Inria) - Hugo Scheithauer (Inria)

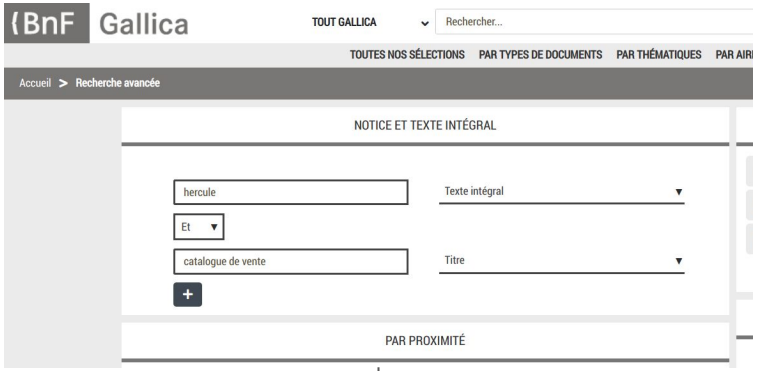
BnF Inria

Institut national d'histoire de l'art INHA

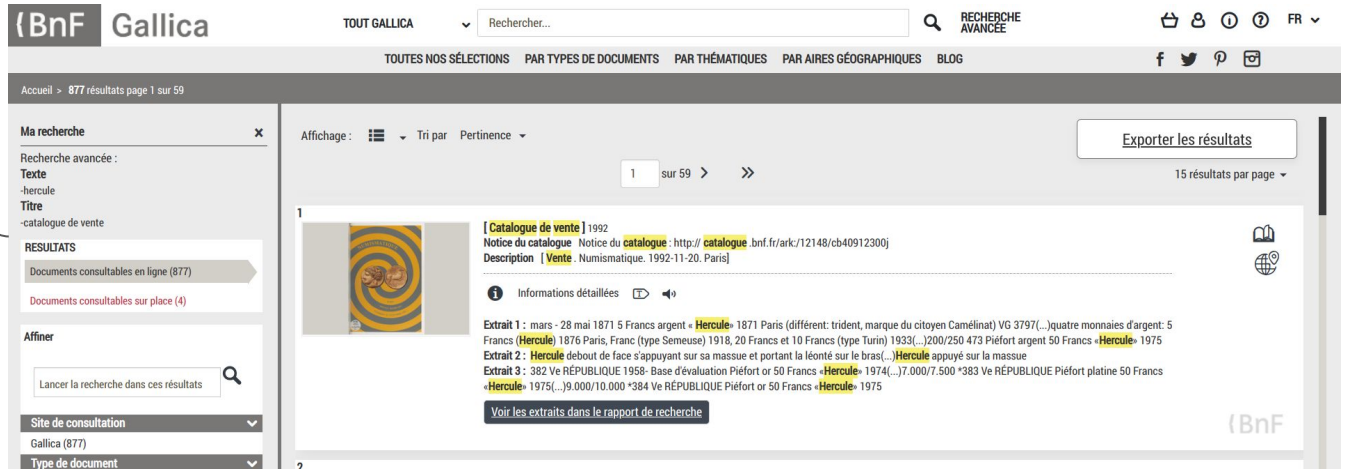

Publication
en ligne
{BnF Gallica
institut
national
d'histoire
de l'art
INHA
Numérisation
Transcription
(OCR...)

Exploration plein texte des catalogues de vente



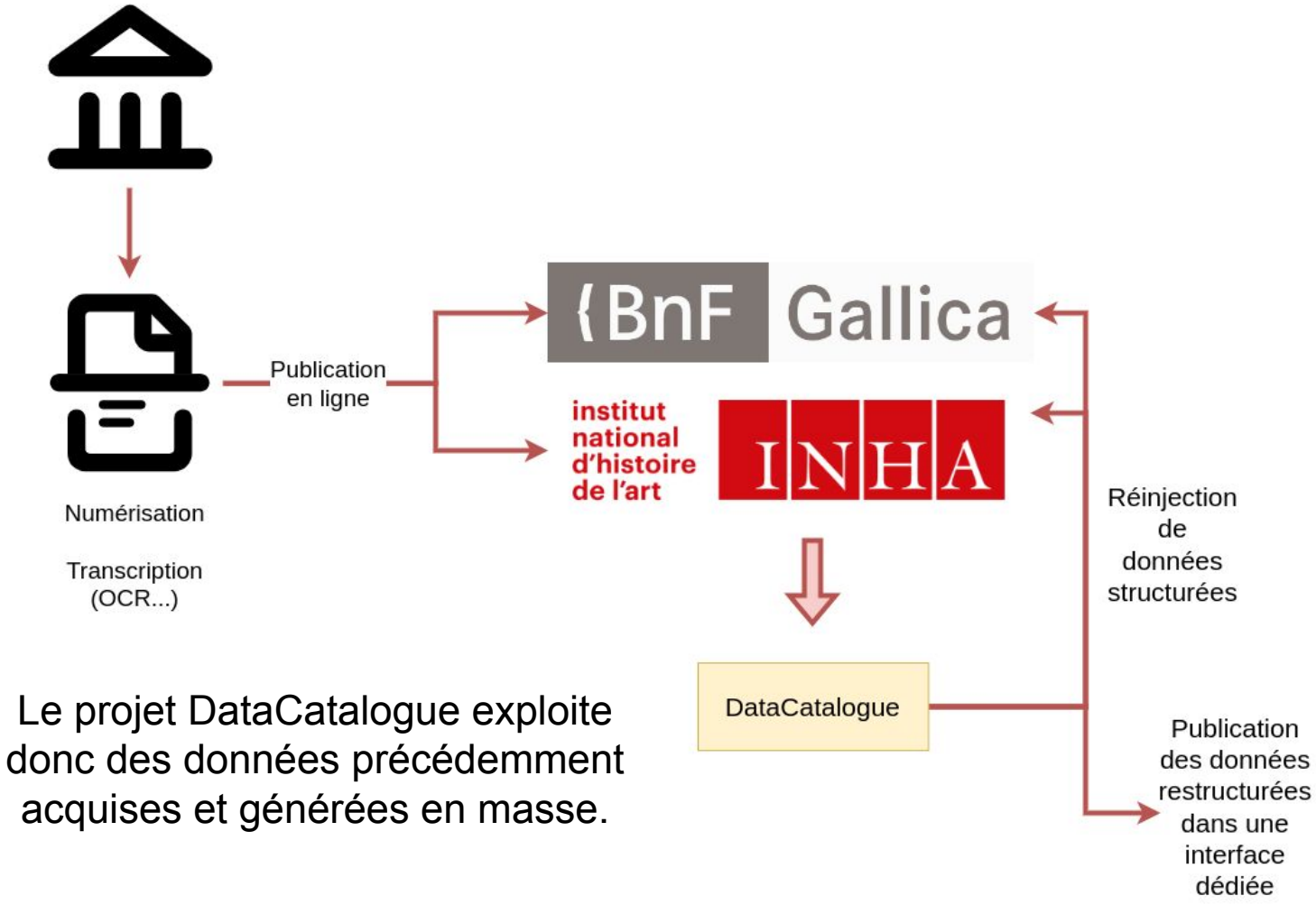


Possibilité d'utiliser des recherches avancées, ici en combinant des termes de recherche.



DataCatalogue : objectifs du projet

- Passer d'une **numérisation** à une **base de données textuelle et requêtable**.
- **Segmenter** les catalogues de vente et attribuer à chaque niveau d'information une **étiquette** : entrées de catalogues, numéro des entrées, description des objets, matériaux, sommes monétaires, etc.
- Produire, à partir des zones segmentées, un **encodage XML-TEI** des catalogues de vente.
- Mettre à disposition des chercheurs le corpus encodé dans une **interface de publication** permettant de **requêter** sur les zones segmentées.



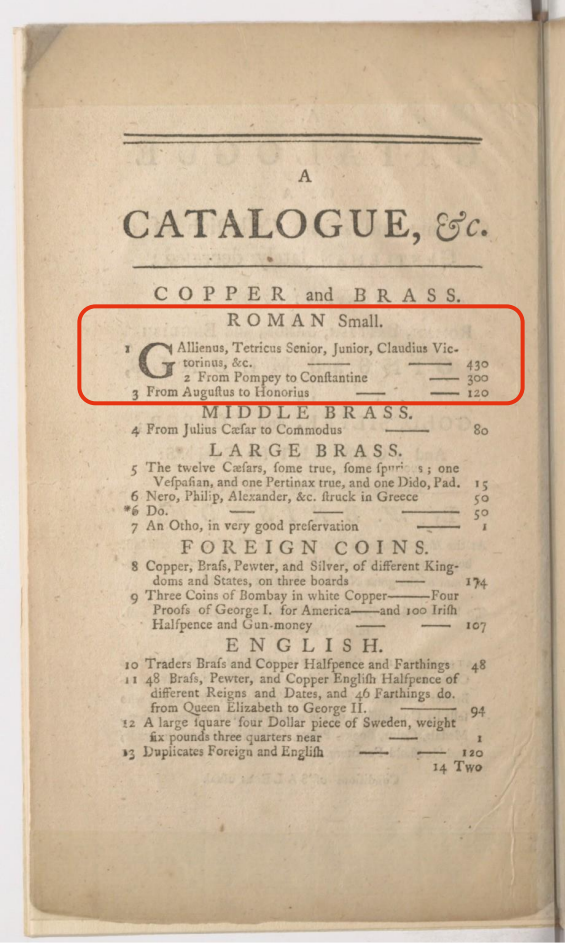
Le projet DataCatalogue exploite donc des données précédemment acquises et générées en masse.

Enjeux

- Création d'un outil permettant de **structurer automatiquement** les catalogues de vente,
- Traiter l'ensemble des catalogues de vente aujourd'hui numérisés, transcrits, et conservés à la BnF et à l'INHA (= trouver une solution généralisable),
- Créer un **nouvel accès** à ces documents pour les publics qui les utilisent.

Approche

- Structuration automatique de l'information,
- Indexation et publication de corpus.



A

CATALOGUE, &c.

COPPER and BRASS.

ROMAN Small.

1 GAllienus, Tetricus Senior, Junior, Claudius Victorinus, &c. —— —— 430
2 From Pompey to Conſtantine —— —— 300
3 From Auguſtus to Honorius —— —— 120

MIDDLE BRASS.

4 From Julius Cæſar to Commodus —— —— 80

LARGE BRASS.

5 The twelve Cæſars, ſome true, ſome ſpuri s; one Veſpaſian, and one Pertinax true, and one Dido, Pad. 15
6 Nero, Philip, Alexander, &c. ſtruck in Greece 50
*6 Do. —— —— —— 50
7 An Otho, in very good preſervation —— —— 1

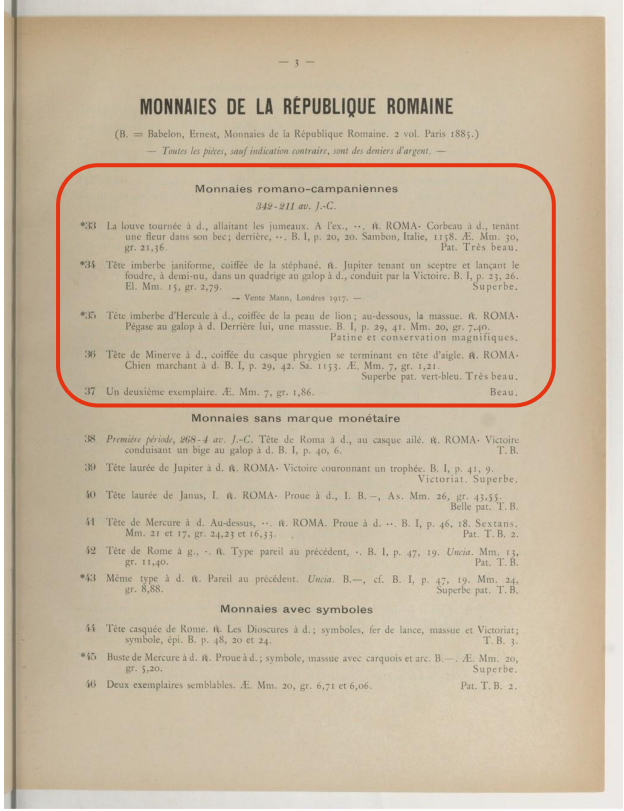
FOREIGN COINS.

8 Copper, Braſs, Pewter, and Silver, of different Kingdoms and States, on three boards —— —— 174
9 Three Coins of Bombay in white Copper——Four Proofs of George I. for America——and 100 Iriſh Halfpence and Gun-money —— —— 107

ENGLISH.

10 Traders Braſs and Copper Halfpence and Farthings 48
11 48 Braſs, Pewter, and Copper Engliſh Halfpence of different Reigns and Dates, and 46 Farthings do. from Queen Elizabeth to George II. —— —— 94
12 A large ſquare four Dollar piece of Sweden, weight ſix pounds three quarters near —— —— 1
13 Duplicates Foreign and Engliſh —— —— 120
14 Two

Whiston Bristow, 1762.



— 3 —

MONNAIES DE LA RÉPUBLIQUE ROMAINE

(B. = Babelon, Ernest, Monnaies de la République Romaine. 2 vol. Paris 1885.)

— Toutes les pièces, sauf indication contraire, sont des deniers d'argent. —

Monnaies romano-campaniennes

348-211 av. J.-C.

*33 La louve tournée à d., allaitant les jumeaux. A l'ex., ··. ℞. ROMA· Corbeau à d., tenant une fleur dans son bec; derrière, ··. B. I, p. 20, 20. Sambon, Italie, 1158. Æ. Mm. 30, gr. 21,36. Pat. Très beau.

*34 Tête imberbe janiforme, coiffée de la stéphané. ℞. Jupiter tenant un sceptre et lançant le foudre, à demi-nu, dans un quadrige au galop à d., conduit par la Victoire. B. I, p. 23, 26. El. Mm. 15, gr. 2,79. Superbe.

→ Vente Mann, Londres 1917. —

*35 Tête imberbe d'Hercule à d., coiffée de la peau de lion; au-dessous, la massue. ℞. ROMA· Pégase au galop à d. Derrière lui, une massue. B. I, p. 29, 41. Mm. 20, gr. 7,40. Patine et conservation magnifiques.

36 Tête de Minerve à d., coiffée du casque phrygien se terminant en tête d'aigle. ℞. ROMA· Chien marchant à d. B. I, p. 29, 42. Sa. 1153. Æ. Mm. 7, gr. 1,21. Superbe pat. vert-bleu. Très beau.

37 Un deuxième exemplaire. Æ. Mm. 7, gr. 1,86. Beau.

Monnaies sans marque monétaire

38 *Première période, 268-4 av. J.-C.* Tête de Roma à d., au casque ailé. ℞. ROMA· Victoire conduisant un bige au galop à d. B. I, p. 40, 6. T. B.

39 Tête laurée de Jupiter à d. ℞. ROMA· Victoire couronnant un trophée. B. I, p. 41, 9. Victoriat. Superbe.

40 Tête laurée de Janus, I. ℞. ROMA· Proue à d., I. B.—, As. Mm. 26, gr. 43,55. Belle pat. T. B.

41 Tête de Mercure à d. Au-dessus, ··. ℞. ROMA. Proue à d. ··. B. I, p. 46, 18. Sextans. Mm. 21 et 17, gr. 24,23 et 16,33. Pat. T. B. 2.

42 Tête de Rome à g., ·. ℞. Type pareil au précédent, ·. B. I, p. 47, 19. *Uncia*. Mm. 13, gr. 11,40. Pat. T. B.

*43 Même type à d. ℞. Pareil au précédent. *Uncia*. B.—, cf. B. I, p. 47, 19. Mm. 24, gr. 8,88. Superbe pat. T. B.

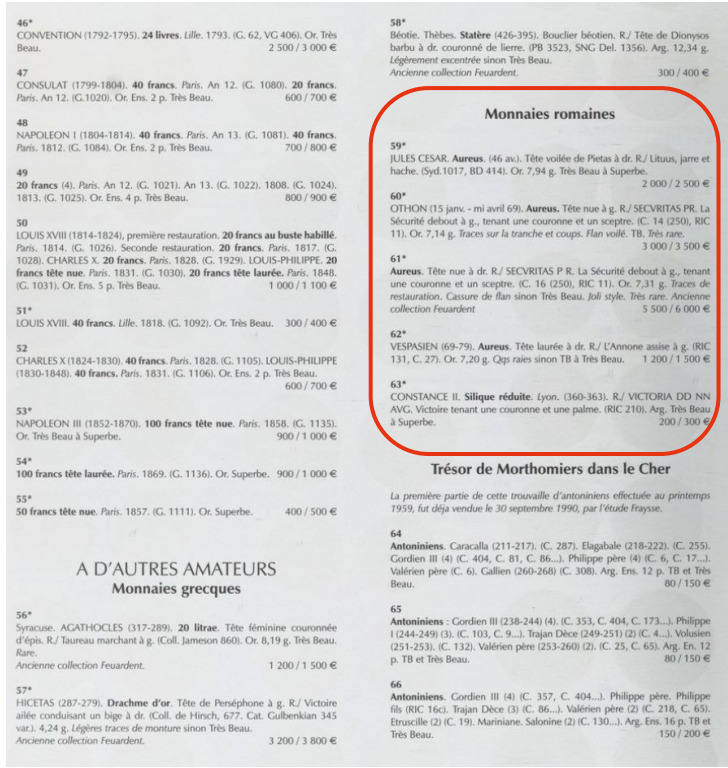
Monnaies avec symboles

44 Tête casquée de Rome. ℞. Les Dioscures à d.; symboles, fer de lance, massue et Victoriat; symbole, épi. B. p. 48, 20 et 24. T. B. 3.

*45 Buste de Mercure à d. ℞. Proue à d.; symbole, massue avec carquois et arc. B.—, Æ. Mm. 20, gr. 5,20. Superbe.

46 Deux exemplaires semblables. Æ. Mm. 20, gr. 6,71 et 6,06. Pat. T. B. 2.

Lucien Naville, 1924.



46*
CONVENTION (1792-1795). **24 livres**. *Lille*. 1793. (G. 62, VG 406). Or. Très Beau. 2 500 / 3 000 €

47
CONSULAT (1799-1804). **40 francs**. *Paris*. An 12. (G. 1080). **20 francs**. *Paris*. An 12. (G.1020). Or. Ens. 2 p. Très Beau. 600 / 700 €

48
NAPOLEON I (1804-1814). **40 francs**. *Paris*. An 13. (G. 1081). **40 francs**. *Paris*. 1812. (G. 1084). Or. Ens. 2 p. Très Beau. 700 / 800 €

49
20 francs (4). *Paris*. An 12. (G. 1021). An 13. (G. 1022). 1808. (G. 1024). 1813. (G. 1025). Or. Ens. 4 p. Très Beau. 800 / 900 €

50
LOUIS XVIII (1814-1824), première restauration. **20 francs au buste habillé**. *Paris*. 1814. (G. 1026). Seconde restauration. **20 francs**. *Paris*. 1817. (G. 1028). CHARLES X. **20 francs**. *Paris*. 1828. (G. 1929). LOUIS-PHILIPPE. **20 francs tête nue**. *Paris*. 1831. (G. 1030). **20 francs tête laurée**. *Paris*. 1848. (G. 1031). Or. Ens. 5 p. Très Beau. 1 000 / 1 100 €

51*
LOUIS XVIII. **40 francs**. *Lille*. 1818. (G. 1092). Or. Très Beau. 300 / 400 €

52
CHARLES X (1824-1830). **40 francs**. *Paris*. 1828. (G. 1105). LOUIS-PHILIPPE (1830-1848). **40 francs**. *Paris*. 1831. (G. 1106). Or. Ens. 2 p. Très Beau. 600 / 700 €

53*
NAPOLEON III (1852-1870). **100 francs tête nue**. *Paris*. 1858. (G. 1135). Or. Très Beau à Superbe. 900 / 1 000 €

54*
100 francs tête laurée. *Paris*. 1869. (G. 1136). Or. Superbe. 900 / 1 000 €

55*
50 francs tête nue. *Paris*. 1857. (G. 1111). Or. Superbe. 400 / 500 €

A D'AUTRES AMATEURS
Monnaies grecques

56*
Syracuse. AGATHOCLES (317-289). **20 litrae**. Tête féminine couronnée d'épis. R./ Taureau marchant à g. (Coll. Jameson 860). Or. 8,19 g. Très Beau. *Rare.*
Ancienne collection Feuardent. 1 200 / 1 500 €

57*
HICETAS (287-279). **Drachme d'or**. Tête de Perséphone à g. R./ Victoire ailée conduisant un bige à dr. (Coll. de Hirsch, 677. Cat. Gulbenkian 345 var.). 4,24 g. *Légères traces de monture* sinon Très Beau.
Ancienne collection Feuardent. 3 200 / 3 800 €

58*
Béotie. Thèbes. **Statère** (426-395). Bouclier béotien. R./ Tête de Dionysos barbu à dr. couronné de lierre. (PB 3523, SNG Del. 1356). Arg. 12,34 g. *Légèrement excentrée* sinon Très Beau.
Ancienne collection Feuardent. 300 / 400 €

Monnaies romaines

59*
JULES CESAR. **Aureus**. (46 av.). Tête voilée de Pietas à dr. R./ Lituus, jarre et hache. (Syd.1017, BD 414). Or. 7,94 g. Très Beau à Superbe. 2 000 / 2 500 €

60*
OTHON (15 janv. - mi avril 69). **Aureus**. Tête nue à g. R./ SECVRITAS PR. La Sécurité debout à g., tenant une couronne et un sceptre. (C. 14 (250), RIC 11). Or. 7,14 g. *Traces sur la tranche et coups. Flan voilé.* TB. *Très rare.* 3 000 / 3 500 €

61*
Aureus. Tête nue à dr. R./ SECVRITAS P R. La Sécurité debout à g., tenant une couronne et un sceptre. (C. 16 (250), RIC 11). Or. 7,31 g. *Traces de restauration. Cassure de flan* sinon Très Beau. *Joli style. Très rare. Ancienne collection Feuardent* 5 500 / 6 000 €

62*
VESPASIEN (69-79). **Aureus**. Tête laurée à dr. R./ L'Annone assise à g. (RIC 131, C. 27). Or. 7,20 g. *Qqs raies* sinon TB à Très Beau. 1 200 / 1 500 €

63*
CONSTANCE II. **Silique réduite**. *Lyon*. (360-363). R./ VICTORIA DD NN AVG. Victoire tenant une couronne et une palme. (RIC 210). Arg. Très Beau à Superbe. 200 / 300 €

Trésor de Morthomiers dans le Cher

La première partie de cette trouvaille d'antoniniens effectuée au printemps 1959, fut déjà vendue le 30 septembre 1990, par l'étude Fraysse.

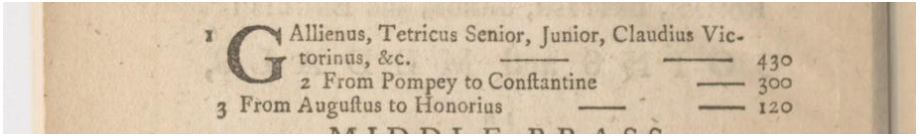
64
Antoniniens. Caracalla (211-217). (C. 287). Elagabale (218-222). (C. 255). Gordien III (4) (C. 404, C. 81, C. 86...). Philippe père (4) (C. 6, C. 17...). Valérien père (C. 6). Gallien (260-268) (C. 308). Arg. Ens. 12 p. TB et Très Beau. 80 / 150 €

65
Antoniniens : Gordien III (238-244) (4). (C. 353, C. 404, C. 173...). Philippe I (244-249) (3). (C. 103, C. 9...). Trajan Dèce (249-251) (2) (C. 4...). Volusien (251-253). (C. 132). Valérien père (253-260) (2). (C. 25, C. 65). Arg. En. 12 p. TB et Très Beau. 80 / 150 €

66
Antoniniens. Gordien III (4) (C. 357, C. 404...). Philippe père. Philippe fils (RIC 16c). Trajan Dèce (3) (C. 86...). Valérien père (2) (C. 218, C. 65). Etruscille (2) (C. 19). Mariniane. Salonine (2) (C. 130...). Arg. Ens. 16 p. TB et Très Beau. 150 / 200 €

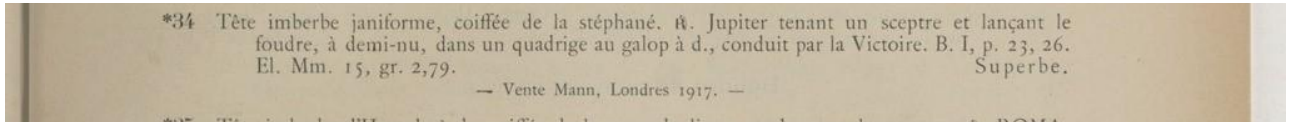
Fraysse & Associés, 2011.

→ Homogénéité dans la structuration de l'information



1 Gallienus, Tetricus Senior, Junior, Claudius Victorinus, &c. —— —— 430
2 From Pompey to Conſtantine —— 300
3 From Auguſtus to Honorius —— —— 120

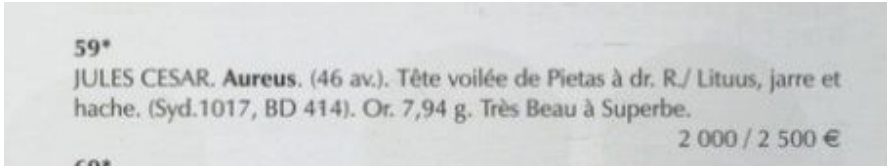
Whiston Bristow, 1762.



*34 Tête imberbe janiforme, coiffée de la stéphané. ℞. Jupiter tenant un sceptre et lançant le foudre, à demi-nu, dans un quadrige au galop à d., conduit par la Victoire. B. I, p. 23, 26. El. Mm. 15, gr. 2,79. Superbe.
— Vente Mann, Londres 1917. —

Lucien Naville, 1924.

→ Mais également une hétérogénéité dans la typographie et la structuration des niveaux d'information atomiques.



59*
JULES CESAR. **Aureus**. (46 av.). Tête voilée de Pietas à dr. R./ Lituus, jarre et hache. (Syd.1017, BD 414). Or. 7,94 g. Très Beau à Superbe.
2 000 / 2 500 €

Fraysse & Associés, 2011.

Quels outils dans notre mallette ?

Une restructuration en XML grâce au standard TEI

- Standard XML pour l'**édition de texte**.
- Permet de rendre **lisible** et **compréhensible** un texte par un ordinateur (*machine-readable*).

```
<p>La TEI met à disposition un ensemble de balises permettant de tagger l'information.</p>
```

Exemple d'encodage TEI du poème *Les Chats*, Baudelaire, 1857.

On donne à la machine les moyens de comprendre du texte comme une base de données.

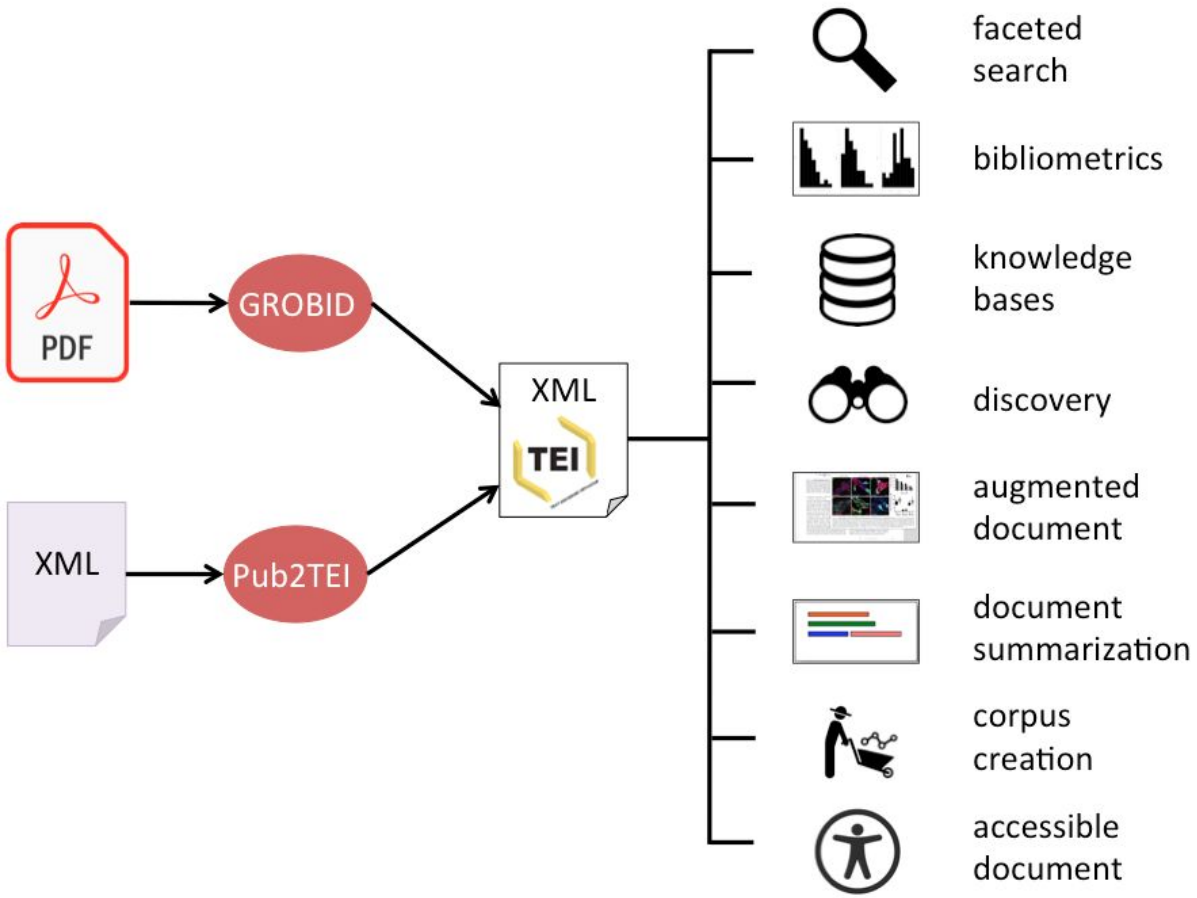
```
<div type="sonnet">
 <lg type="quatrain">
  <l>Les amoureux fervents et les savants austères</l>
  <l> Aiment également, dans leur mûre saison,</l>
  <l> Les chats puissants et doux, orgueil de la maison,</l>
  <l> Qui comme eux sont frileux et comme eux sédentaires.</l>
 </lg>
 <lg type="quatrain">
  <l>Amis de la science et de la volupté</l>
  <l> Ils cherchent le silence et l'horreur des ténèbres ;</l>
  <l> L'Erèbe les eût pris pour ses coursiers funèbres,</l>
  <l> S'ils pouvaient au servage incliner leur fierté.</l>
 </lg>
 <lg type="tercet">
  <l>Ils prennent en songeant les nobles attitudes</l>
  <l>Des grands sphinx allongés au fond des solitudes,</l>
  <l>Qui semblent s'endormir dans un rêve sans fin ;</l>
 </lg>
 <lg type="tercet">
  <l>Leurs reins féconds sont pleins d'étincelles magiques,</l>
  <l> Et des parcelles d'or, ainsi qu'un sable fin,</l>
  <l>Etoilent vaguement leurs prunelles mystiques.</l>
 </lg>
</div>
```

GROBID (GeneRation Of BIbliographic Data)

https://github.com/kermitt2/grobid

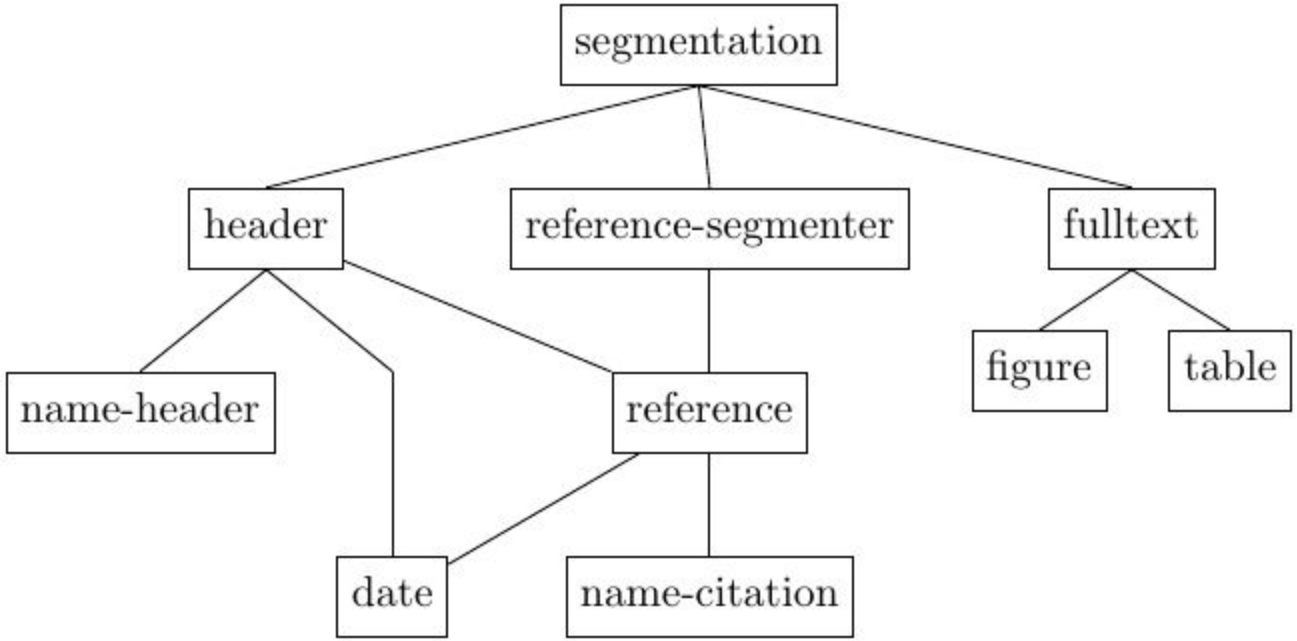
- Librairie d'apprentissage machine pour **extraire**, **parser** et **re-structurer** des documents **PDF** en **XML-TEI**, créée par Patrice Lopez en 2008.
- Objectif : avoir un outil pour **structurer automatiquement** les publications scientifiques.
- Outil ***open source***.
- Déployé sur plusieurs plateformes : ResearchGate, HAL, INIST-CNRS, *etc.*

Sur HAL, par exemple, GROBID permet de d'extraire automatiquement les métadonnées et les références d'une publication.

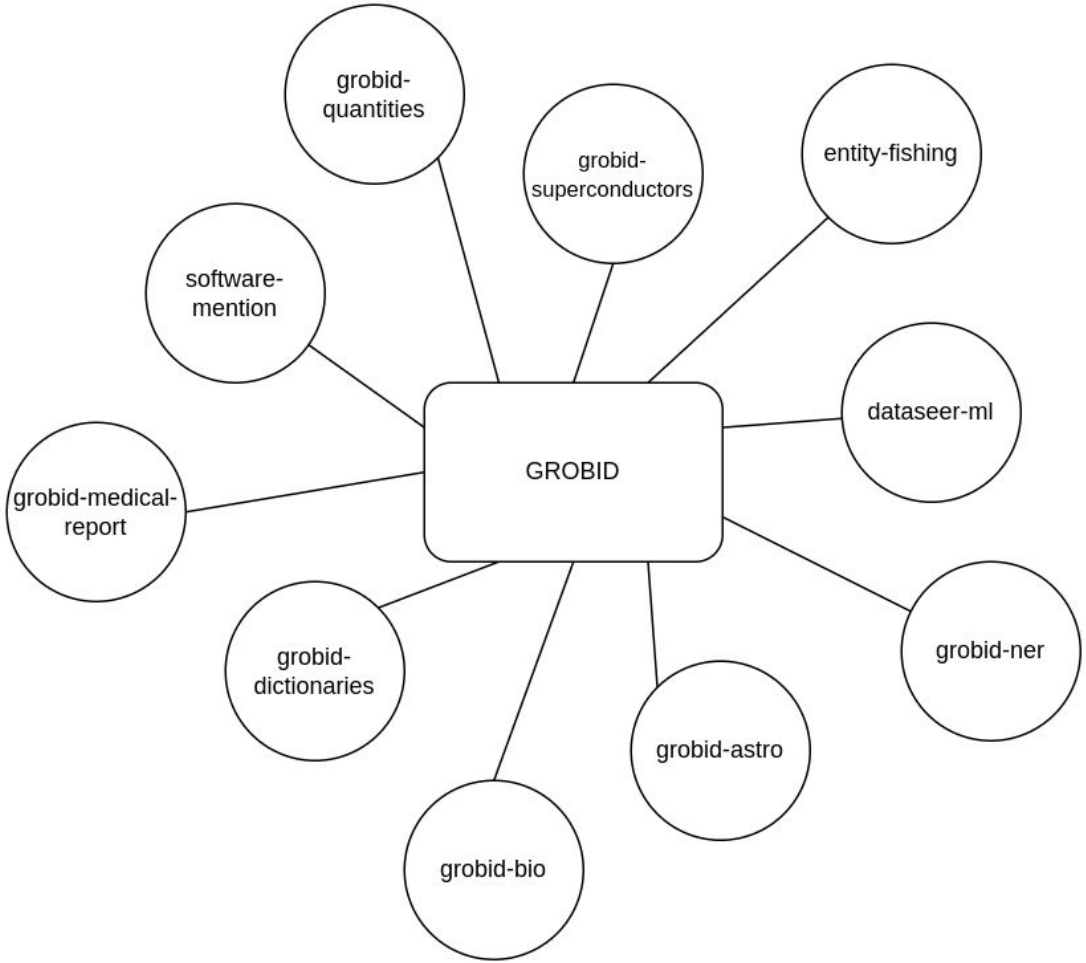


Source : https://grobid.readthedocs.io/en/latest/Principles/

GROBID utilise une cascade de modèles CRF (Conditional Random Fields / Champs aléatoires conditionnels) :

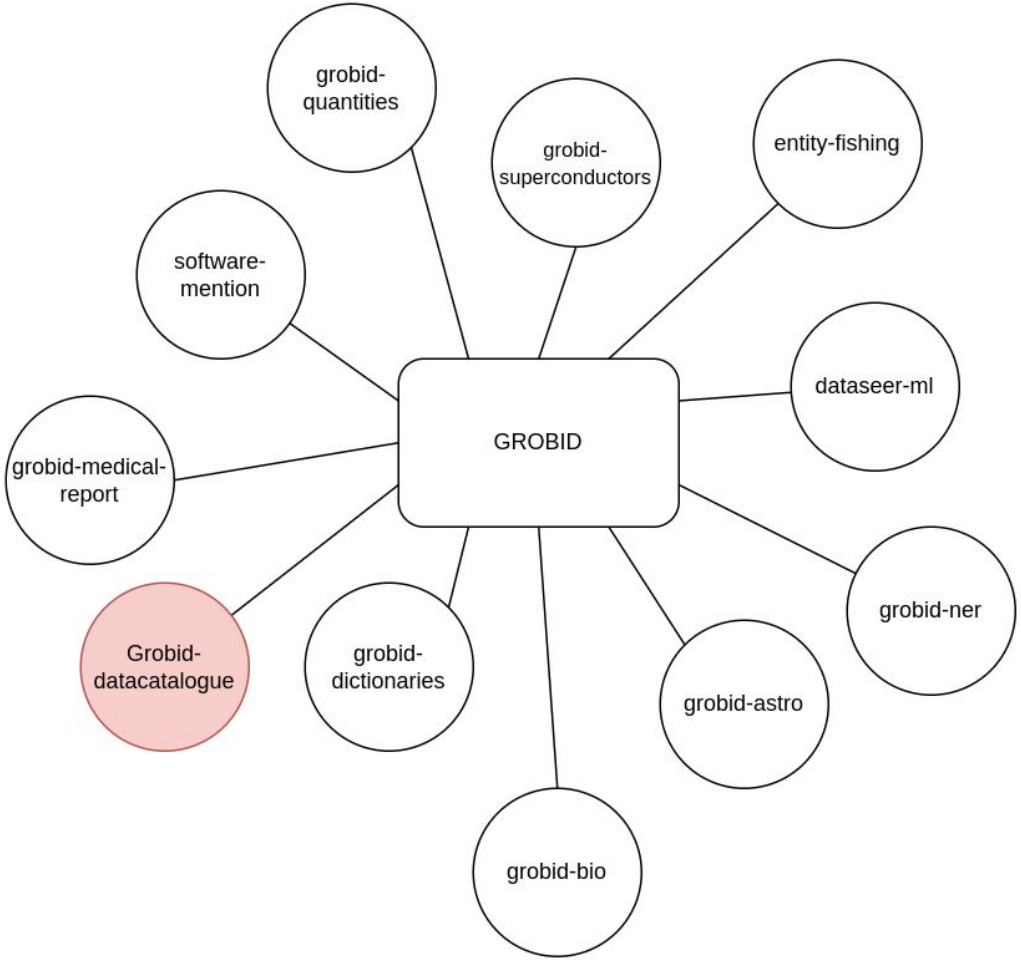


Source : https://grobid.readthedocs.io/en/latest/Principles/



GROBID est aujourd'hui accompagné de plusieurs **modules**.

Un module est un élément dépendant de GROBID, mais spécialisé sur un autre type de document.



L'objectif est de créer un module GROBID dédié au format que représentent les catalogues de vente.

Publier les contenus restructurés avec TEI Publisher

- Application de **publication de fichier TEI**,
- *Open source*,
- Permet d'**indexer** un corpus, de le **visualiser**, et de le rendre **requêtable**,
- Interface entièrement **personnalisable**.

Ingénierie de projet

- **Modélisation TEI des catalogues de vente**

→ Comment modéliser la structure et les informations contenues dans les catalogues de vente ?

→ Quelle granularité vise-t-on ?

- **Développement du module GROBID DataCatalogue**

→ Ingénierie logicielle sur GROBID,

→ Sciences des données,

→ Constitution des corpus d'entraînement, entraînement et évaluation des modèles,

→ Mise en place de la chaîne de traitement.

- **Publication des catalogues encodés**

→ Quelle stratégie d'éditorialisation ?

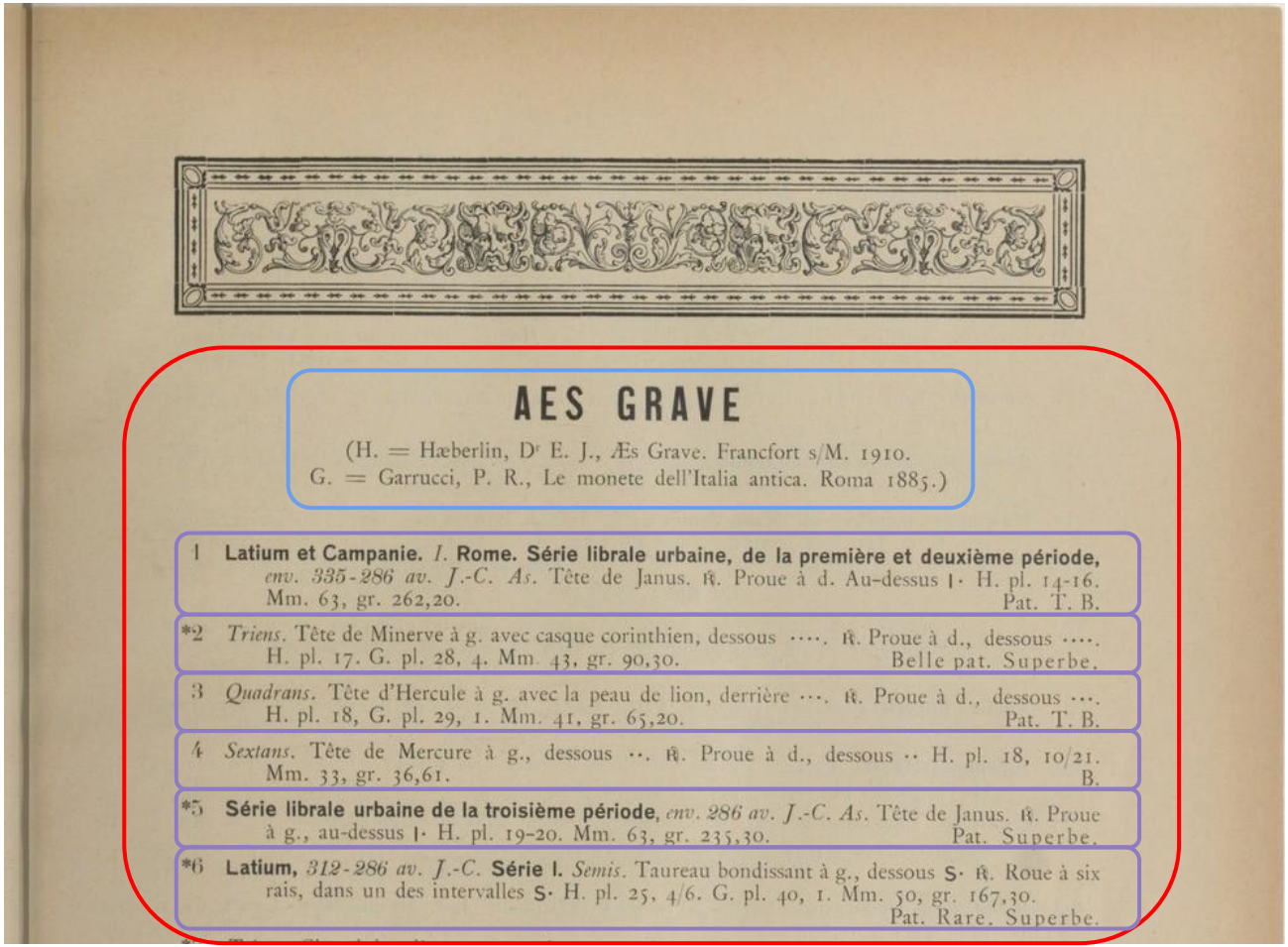
→ Quelles fonctionnalités ?

Modéliser les catalogues de vente avec le standard XML-TEI

La TEI offre la possibilité de créer des **schémas personnalisés** selon les besoins des projets qui l'utilisent, tout en restant conforme au standard.

Pour DataCatalogue, nous avons mis au point un **schéma d'encodage** des catalogues de vente. Celui-ci s'appuie sur :

- Les éléments mis à disposition par la TEI, notamment pour la **description d'objets**.
- Des **éléments personnalisés** permettant de rendre compte de la structuration de l'information des catalogues.



AES GRAVE

(H. = Hæberlin, Dr E. J., Æs Grave. Francfort s/M. 1910.
G. = Garrucci, P. R., Le monete dell'Italia antica. Roma 1885.)

1 **Latium et Campanie.** *I.* **Rome. Série librale urbaine, de la première et deuxième période,** *env. 335-286 av. J.-C. As.* Tête de Janus. ℞. Proue à d. Au-dessus I· H. pl. 14-16. Mm. 63, gr. 262,20. Pat. T. B.

*2 *Triens.* Tête de Minerve à g. avec casque corinthien, dessous ····. ℞. Proue à d., dessous ····. H. pl. 17. G. pl. 28, 4. Mm. 43, gr. 90,30. Belle pat. Superbe.

3 *Quadrans.* Tête d'Hercule à g. avec la peau de lion, derrière ···. ℞. Proue à d., dessous ···. H. pl. 18, G. pl. 29, 1. Mm. 41, gr. 65,20. Pat. T. B.

4 *Sextans.* Tête de Mercure à g., dessous ··. ℞. Proue à d., dessous ·· H. pl. 18, 10/21. Mm. 33, gr. 36,61. B.

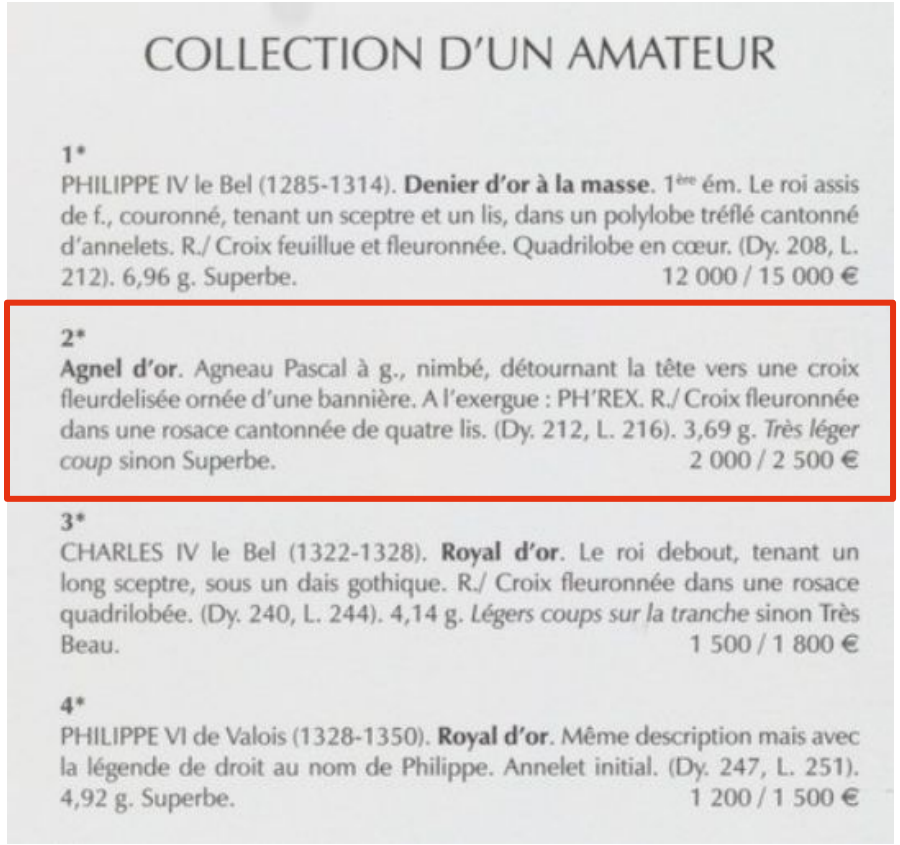
*5 **Série librale urbaine de la troisième période,** *env. 286 av. J.-C. As.* Tête de Janus. ℞. Proue à g., au-dessus I· H. pl. 19-20. Mm. 63, gr. 235,30. Pat. Superbe.

*6 **Latium,** *312-286 av. J.-C.* **Série I.** *Semis.* Taureau bondissant à g., dessous S· ℞. Roue à six rais, dans un des intervalles S· H. pl. 25, 4/6. G. pl. 40, 1. Mm. 50, gr. 167,30. Pat. Rare. Superbe.

- Entrée de catalogue
- Niveaux de titre et descriptions
- Notices

- **catalogueEntry** : Ensemble cohérent de notices d'objets mis en vente et de leurs métadonnées (peut être récursif).
- **catalogueDesc** : métadonnées relatives aux notices (typologie, datation, informations bibliographiques, nom de la collection, des collectionneurs, etc.)
- **catalogueItem** : notice décrivant un objet mis en vente + injection du module TEI *msdescription*.

Voir également : Simon Gabay, Barbara Topalov, Caroline Corbières, Lucie Rondeau Du Noyer, Béatrice Joyeux-Prunel, et al.. Automating Artl@s – extracting data from exhibition catalogues. EADH 2021 - *Second International Conference of the European Association for Digital Humanities*, Sep 2021, Krasnoyarsk, Russia. ⟨hal-03331838⟩



COLLECTION D'UN AMATEUR

1*
PHILIPPE IV le Bel (1285-1314). **Denier d'or à la masse**. 1ère ém. Le roi assis de f., couronné, tenant un sceptre et un lis, dans un polylobe tréflé cantonné d'annelets. R./ Croix feuillue et fleuronnée. Quadrilobe en cœur. (Dy. 208, L. 212). 6,96 g. Superbe. 12 000 / 15 000 €

2*
Agnel d'or. Agneau Pascal à g., nimbé, détournant la tête vers une croix fleurdelisée ornée d'une bannière. A l'exergue : PH'REX. R./ Croix fleuronnée dans une rosace cantonnée de quatre lis. (Dy. 212, L. 216). 3,69 g. *Très léger coup* sinon Superbe. 2 000 / 2 500 €

3*
CHARLES IV le Bel (1322-1328). **Royal d'or**. Le roi debout, tenant un long sceptre, sous un dais gothique. R./ Croix fleuronnée dans une rosace quadrilobée. (Dy. 240, L. 244). 4,14 g. *Légers coups sur la tranche* sinon Très Beau. 1 500 / 1 800 €

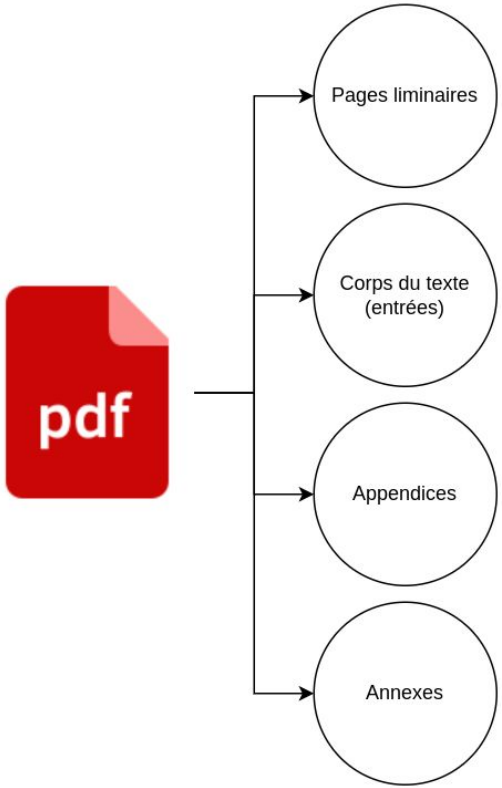
4*
PHILIPPE VI de Valois (1328-1350). **Royal d'or**. Même description mais avec la légende de droit au nom de Philippe. Annelet initial. (Dy. 247, L. 251). 4,92 g. Superbe. 1 200 / 1 500 €

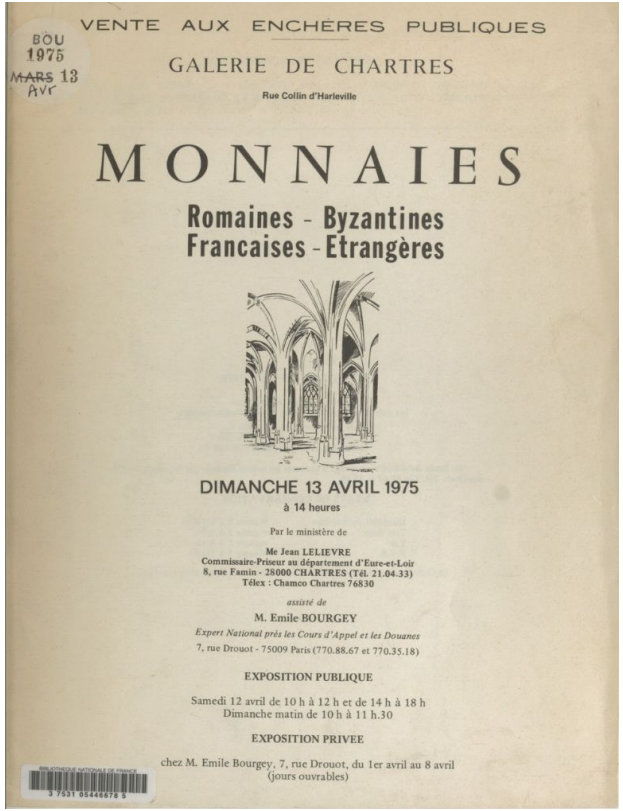
Collection d'un amateur - A d'autres amateurs / Fraysse & Associés ; Sabine Bourgey, 2011, p. 2 (extrait).

```xml
<catalogueEntry>
    <catalogueDesc>
        <head>Collection d'un amateur</head>
    </catalogueDesc>
    <!-- ... -->
    <catalogueItem>
        <altIdentifier>
            <idno>2</idno>
        </altIdentifier>
        <metamark>*</metamark>
        <objectDesc>
            <supportDesc>
                <support>Agnel d'or.</support>
            </supportDesc>
        </objectDesc>
        <decoDesc>
            <ab>Agneau Pascal à g., nimbé, détournant la tête vers une
                croix fleurdelisée ornée d'une bannière. A l'exergue:
                PH'REX R./ Croix fleuronnée dans une rosace cantonnée de quatre
                lis.</ab>
        </decoDesc>
        <objectDesc>
            <supportDesc>
                <support>(<measure>Dy. 212</measure>,
                    <measure>L. 216</measure>).
                    <measure>3,69 g.</measure></support>
                <condition>Très léger coup sinon Superbe<p>.</p></condition>
            </supportDesc>
        </objectDesc>
        <num type="currency">2000 / 2500 euros</num>
    </catalogueItem>
    <!-- ... -->
</catalogueEntry>
```

Développement du module GROBID DataCatalogue

Premier niveau de segmentation : segmentation de haut niveau





VENTE AUX ENCHÈRES PUBLIQUES

GALERIE DE CHARTRES

Rue Collin d'Harleville

MONNAIES

Romaines - Byzantines
Francaises - Etrangères

DIMANCHE 13 AVRIL 1975

à 14 heures

Par le ministère de

Me Jean LELIEVRE
Commissaire-Priseur au département d'Eure-et-Loir
8, rue Famin - 28000 CHARTRES (Tél. 21.04.33)
Télex : Chamco Chartres 76830

assisté de

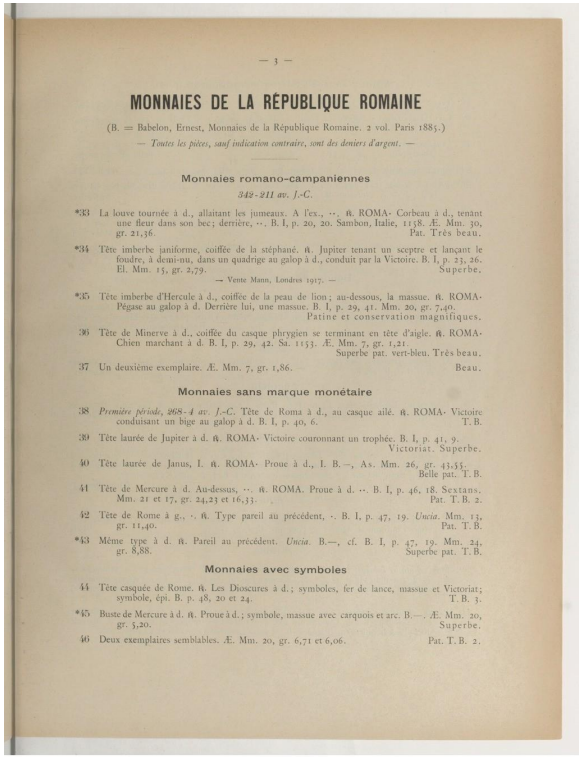
M. Emile BOURGEY
Expert National près les Cours d'Appel et les Douanes
7, rue Drouot - 75009 Paris (770.88.67 et 770.35.18)

EXPOSITION PUBLIQUE

Samedi 12 avril de 10 h à 12 h et de 14 h à 18 h
Dimanche matin de 10 h à 11 h.30

EXPOSITION PRIVEE

chez M. Emile Bourgey, 7, rue Drouot, du 1er avril au 8 avril
(jours ouvrables)



MONNAIES DE LA RÉPUBLIQUE ROMAINE

(B. = Babelon, Ernest, Monnaies de la République Romaine. 2 vol. Paris 1885.)

— Toutes les pièces, sauf indication contraire, sont des deniers d'argent. —

Monnaies romano-campaniennes

342-211 av. J.-C.

*33 La louve tournée à d., allaitant les jumeaux. A l'ex., ··. ℞. ROMA· Corbeau à d., tenant une fleur dans son bec; derrière, ··. B. I, p. 20, 20. Sambon, Italie, 1158. Æ. Mm. 30, gr. 21,36. Pat. Très beau.

*34 Tête imberbe janiforme, coiffée de la stéphané. ℞. Jupiter tenant un sceptre et lançant le foudre, à demi-nu, dans un quadrige au galop à d., conduit par la Victoire. B. I, p. 23, 26. El. Mm. 15, gr. 2,79. Superbe.

→ Vente Mann, Londres 1917. —

*35 Tête imberbe d'Hercule à d., coiffée de la peau de lion; au-dessous, la massue. ℞. ROMA· Pégase au galop à d. Derrière lui, une massue. B. I, p. 29, 41. Mm. 20, gr. 7,40. Patine et conservation magnifiques.

36 Tête de Minerve à d., coiffée du casque phrygien se terminant en tête d'aigle. ℞. ROMA· Chien marchant à d. B. I, p. 29, 42. Sa. 1153. Æ. Mm. 7, gr. 1,21. Superbe pat. vert-bleu. Très beau.

37 Un deuxième exemplaire. Æ. Mm. 7, gr. 1,86. Beau.

Monnaies sans marque monétaire

38 *Première période, 268-4 av. J.-C.* Tête de Roma à d., au casque ailé. ℞. ROMA· Victoire conduisant un bige au galop à d. B. I, p. 40, 6. T. B.

39 Tête laurée de Jupiter à d. ℞. ROMA· Victoire couronnant un trophée. B. I, p. 41, 9. Victoriat. Superbe.

40 Tête laurée de Janus, I. ℞. ROMA· Proue à d., I. B. —, As. Mm. 26, gr. 43,55. Belle pat. T. B.

41 Tête de Mercure à d. Au-dessus, ··. ℞. ROMA. Proue à d. ··. B. I, p. 46, 18. Sextans. Mm. 21 et 17, gr. 24,23 et 16,33. Pat. T. B. 2.

42 Tête de Rome à g., ·. ℞. Type pareil au précédent, ·. B. I, p. 47, 19. *Uncia*. Mm. 13, gr. 11,40. Pat. T. B.

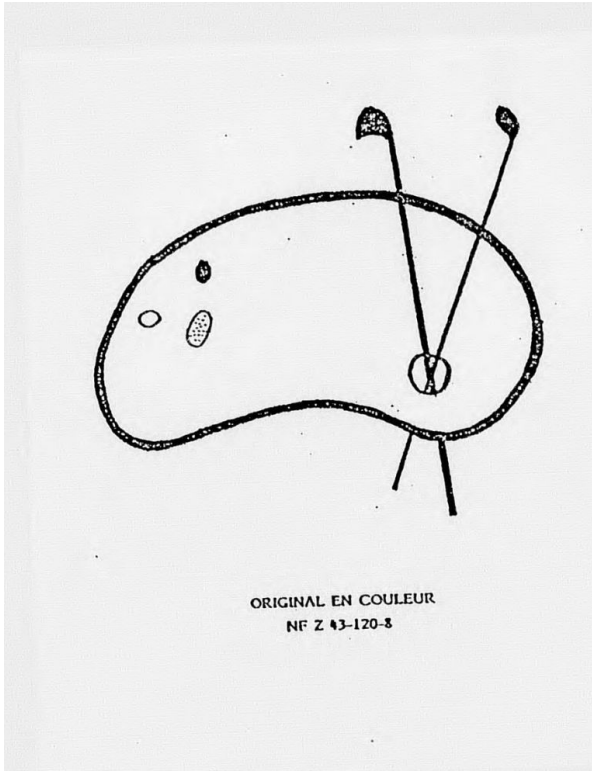
*43 Même type à d. ℞. Pareil au précédent. *Uncia*. B.—, cf. B. I, p. 47, 19. Mm. 24, gr. 8,88. Superbe pat. T. B.

Monnaies avec symboles

44 Tête casquée de Rome. ℞. Les Dioscures à d.; symboles, fer de lance, massue et Victoriat; symbole, épi. B. p. 48, 20 et 24. T. B. 3.

*45 Buste de Mercure à d. ℞. Proue à d.; symbole, massue avec carquois et arc. B.—. Æ. Mm. 20, gr. 5,20. Superbe.

46 Deux exemplaires semblables. Æ. Mm. 20, gr. 6,71 et 6,06. Pat. T. B. 2.



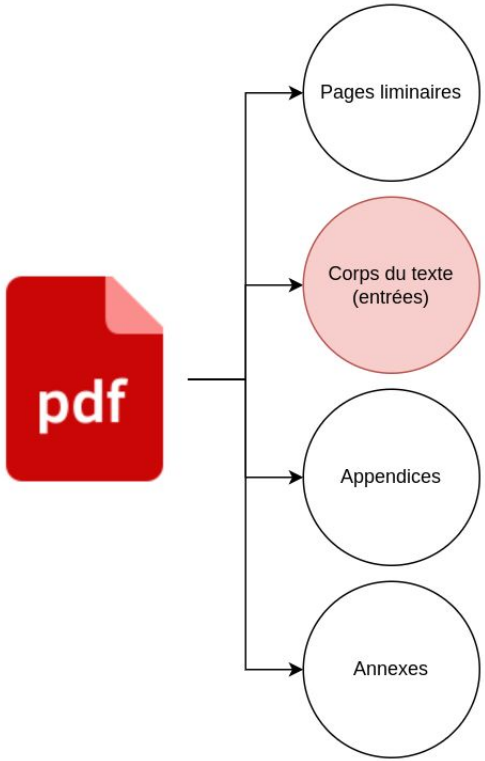
ORIGINAL EN COULEUR
NF Z 43-120-8

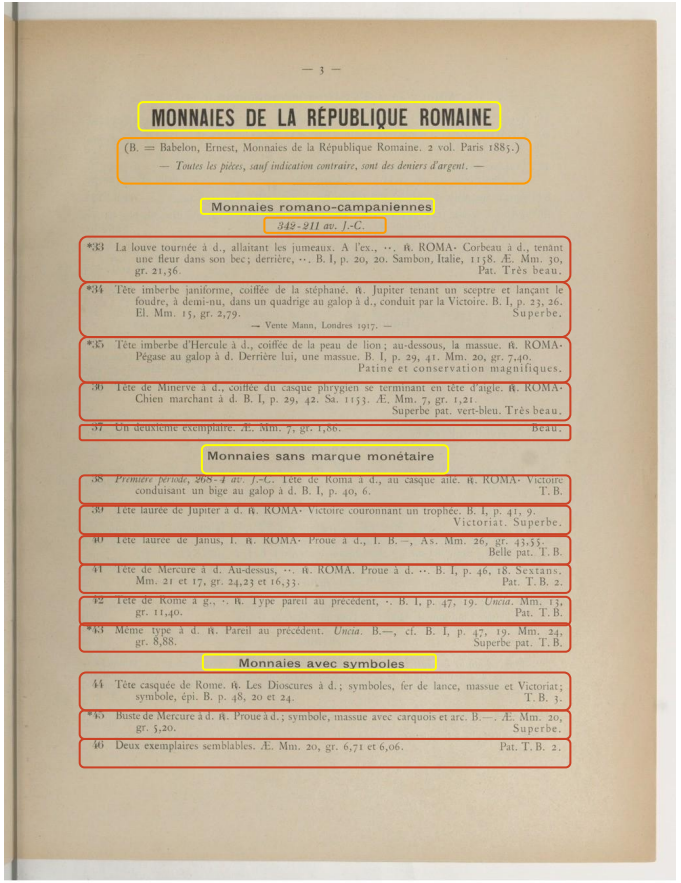
Éléments liminaires

Corps du texte

Appendices

Deuxième niveau : segmentation des entrées du catalogue





— 3 —

MONNAIES DE LA RÉPUBLIQUE ROMAINE

(B. = Babelon, Ernest, Monnaies de la République Romaine. 2 vol. Paris 1885.)

— Toutes les pièces, sauf indication contraire, sont des deniers d'argent. —

Monnaies romano-campaniennes

349-211 av. J.-C.

**13* La louve tournée à d., allaitant les jumeaux. A l'ex., ··. ℞. ROMA· Corbeau à d., tenant une fleur dans son bec; derrière, ··. B. I, p. 20, 20. Sambon, Italie, 1158. Æ. Mm. 30, gr. 21,36. Pat. Très beau.

**14* Tête imberbe janiforme, coiffée de la stéphané. ℞. Jupiter tenant un sceptre et lançant le foudre, à demi-nu, dans un quadrige au galop à d., conduit par la Victoire. B. I, p. 23, 26. El. Mm. 15, gr. 2,79. Superbe.

→ Vente Mann, Londres 1917. —

**15* Tête imberbe d'Hercule à d., coiffée de la peau de lion; au-dessous, la massue. ℞. ROMA· Pégase au galop à d. Derrière lui, une massue. B. I, p. 29, 41. Mm. 20, gr. 7,40. Patine et conservation magnifiques.

16 Tête de Minerve à d., coiffée du casque phrygien se terminant en tête d'aigle. ℞. ROMA· Chien marchant à d. B. I, p. 29, 42. Sa. 1153. Æ. Mm. 7, gr. 1,21. Superbe pat. vert-bleu. Très beau.

17 Un deuxième exemplaire. Æ. Mm. 7, gr. 1,86. Beau.

Monnaies sans marque monétaire

18 *Première période, 268-4 av. J.-C.* Tête de Roma à d., au casque ailé. ℞. ROMA· Victoire conduisant un bige au galop à d. B. I, p. 40, 6. T. B.

19 Tête laurée de Jupiter à d. ℞. ROMA· Victoire couronnant un trophée. B. I, p. 41, 9. Victoriat. Superbe.

20 Tête laurée de Janus, I. ℞. ROMA· Proue à d., I. B.—, As. Mm. 26, gr. 43,55. Belle pat. T. B.

21 Tête de Mercure à d. Au-dessus, ··. ℞. ROMA. Proue à d. ··. B. I, p. 46, 18. Sextans. Mm. 21 et 17, gr. 24,23 et 16,33. Pat. T. B. 2.

22 Tête de Rome à g., ·. ℞. Type pareil au précédent, ·. B. I, p. 47, 19. *Uncia*. Mm. 13, gr. 11,40. Pat. T. B.

**23* Même type à d. ℞. Pareil au précédent. *Uncia*. B.—, cf. B. I, p. 47, 19. Mm. 24, gr. 8,88. Superbe pat. T. B.

Monnaies avec symboles

24 Tête casquée de Rome. ℞. Les Dioscures à d.; symboles, fer de lance, massue et Victoriat; symbole, épi. B. p. 48, 20 et 24. T. B. 3.

**25* Buste de Mercure à d. ℞. Proue à d.; symbole, massue avec carquois et arc. B.—. Æ. Mm. 20, gr. 5,20. Superbe.

26 Deux exemplaires semblables. Æ. Mm. 20, gr. 6,71 et 6,06. Pat. T. B. 2.

Source gallica.bnf.fr / Bibliothèque nationale de France

Troisième niveau et suivants : segmentation du contenu des notices des catalogues

*35 Tête imberbe d'Hercule à d., coiffée de la peau de lion ; au-dessous, la massue. ℞. ROMA· Pégase au galop à d. Derrière lui, une massue. B. I, p. 29, 41. Mm. 20, gr. 7,40. Patine et conservation magnifiques.

*35

Tête imberbe d'Hercule à d., coiffée de la peau de lion ; au-dessous, la massue. ℞. ROMA· Pégase au galop à d. Derrière lui, une massue.

Plusieurs éléments à segmenter ici : numéros d'items ; descriptions ; entités nommées ; éléments bibliographiques ; quantités/mesures ; observation de conservations.

B. I, p. 29, 41.

Mm. 20, gr. 7,40.

Patine et conservation magnifiques.

Création d’un corpus pour l’entraînement des modèles GROBID DataCatalogue

BnF :

- Bienaimé-Feuardent
- Bourgey
- Naville
- XVIIIe siècle

INHA :

- Desvouges
- Lair-Dubreuil

Catalogues propriétaires :

- CGB
- CNG

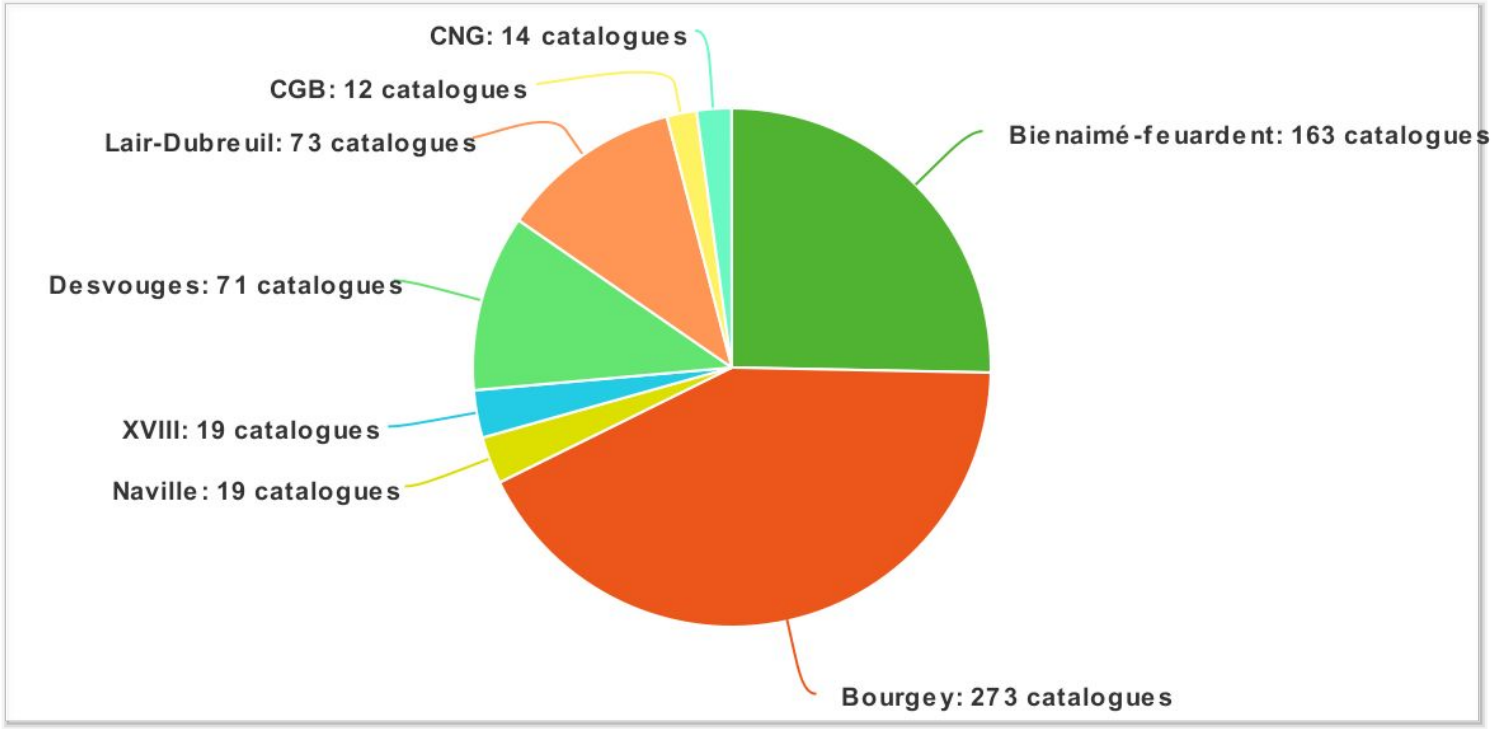
Cet échantillon a pour objectif d’être représentatif en termes de datation, de type de vente, de typographie et typologie.

Siècles représentés : XVIIIe, XIXe, XXe, XXIe.

Langues : ~95% en français.

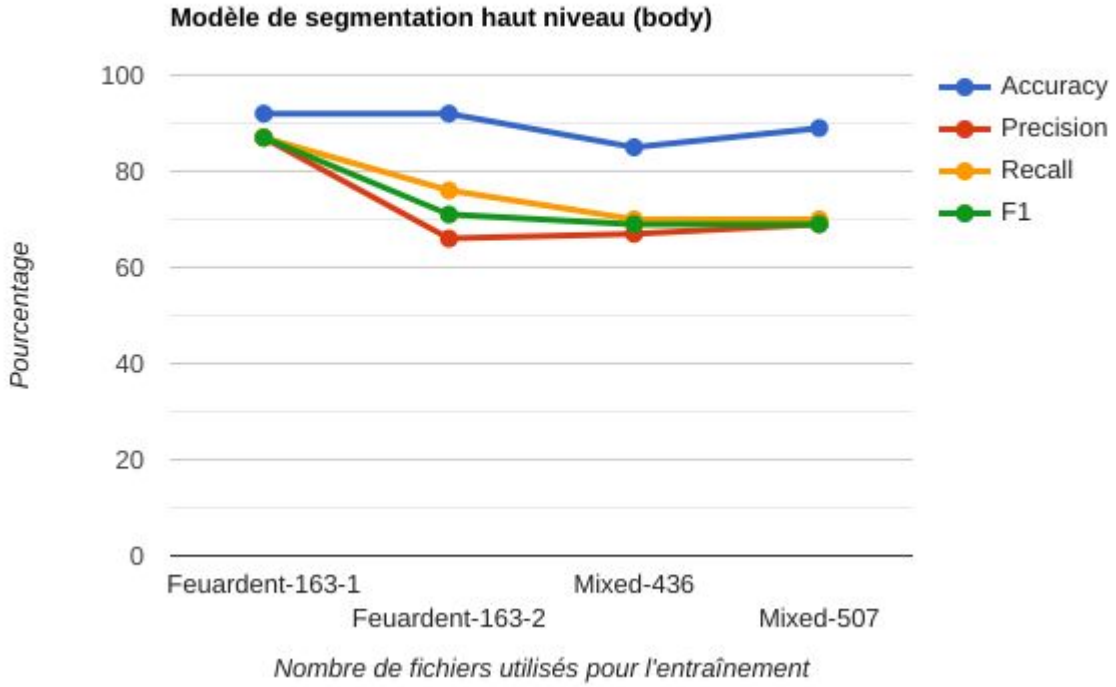
Types de vente : numismatique, livres, antiquités, objets d’art, objets de luxe (catalogues propriétaires).

RÉPARTITION DU CORPUS



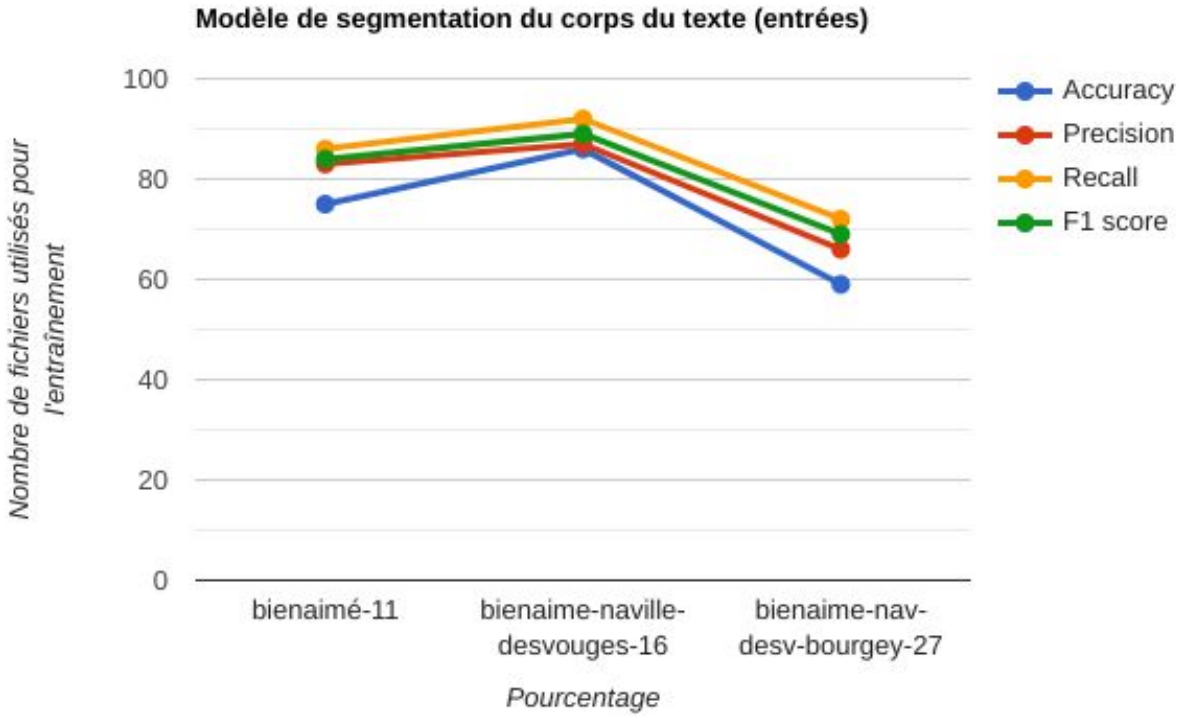
Évaluer les performances des modèles GROBID

Quelles performances pour le modèle de segmentation haut niveau ?



- Mixed-436 : Bienaimé-Feuardent + Bourgey
- Mixed-507 : Bienaimé-Feuardent + Bourgey + Desvouges

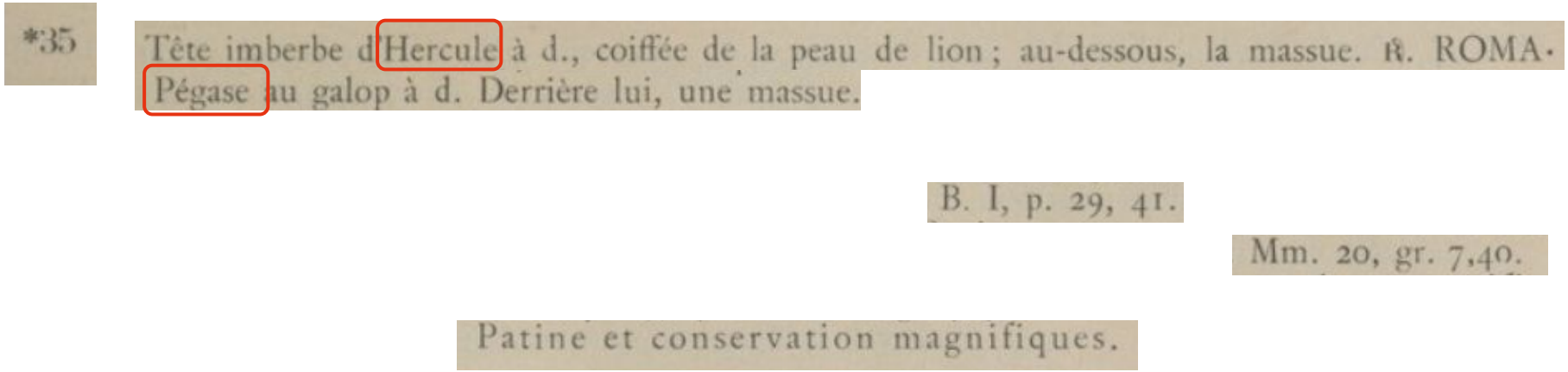
Quelles performances pour la segmentation des entrées de catalogue ?



- Scores en hausse pour l'entraînement d'un modèle sur les collections Bienaimé-feuardent, Naville et Desvouges, mais réserve car le nombre de fichiers doit être augmenté.

- L'utilisation de la collection Bourgey dans l'entraînement vient détériorer les scores.

Viser un encodage fin de l'information dans les entrées de catalogues : la reconnaissance d'entités nommées



*35

Tête imberbe d'Hercule à d., coiffée de la peau de lion ; au-dessous, la massue. ℞. ROMA· Pégase au galop à d. Derrière lui, une massue.

B. I, p. 29, 41.

Mm. 20, gr. 7,40.

Patine et conservation magnifiques.

Exemple : détection automatique des noms de personnages mythologiques

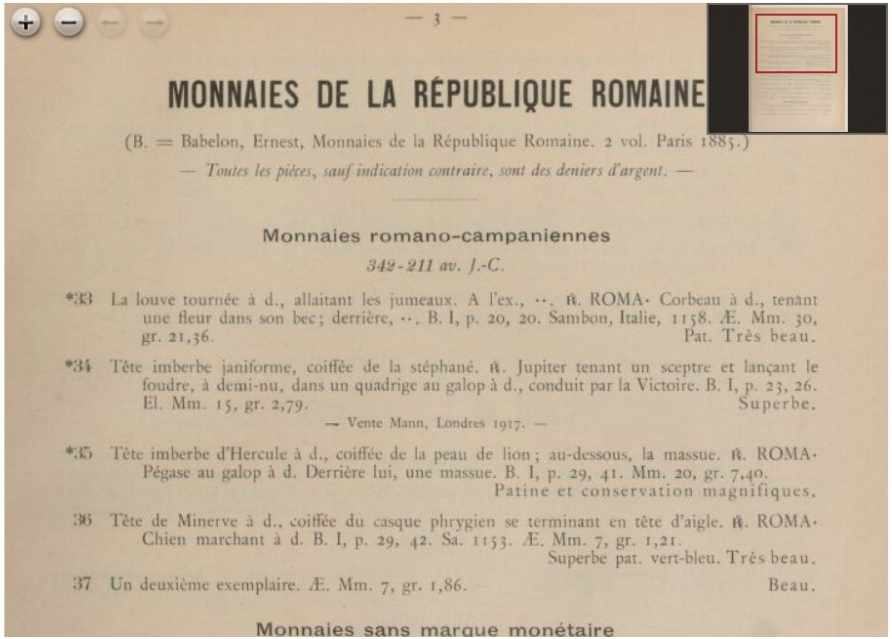
Stage de fin d’études réalisé par Abderraouf Farhi (Université de Tours, M2 Intelligence des données de la culture et des patrimoines)

Extraction automatique d’information dans les catalogues de vente

- **Exploration des outils d’extraction d’information** déjà développés, autres que GROBID, et disponible librement en ligne (principalement SpaCy),
- Test de **modèles pré-entraînés de reconnaissance d’entités nommées** sur le corpus des catalogues de vente,
- Proposition d’intégration d’une étape de reconnaissance d’entités nommées dans la chaîne de traitement DataCatalogue.

Publier les catalogues de vente encodés

AJOUTER UNE VUE



— 3 —

MONNAIES DE LA RÉPUBLIQUE ROMAINE

(B. = Babelon, Ernest, Monnaies de la République Romaine. 2 vol. Paris 1885.)

— *Toutes les pièces, sauf indication contraire, sont des deniers d'argent.* —

Monnaies romano-campaniennes

342-211 av. J.-C.

*33 La louve tournée à d., allaitant les jumeaux. A l'ex., ··. ℞. ROMA· Corbeau à d., tenant une fleur dans son bec; derrière, ··. B. I, p. 20, 20. Sambon, Italie, 1158. Æ. Mm. 30, gr. 21,36. Pat. Très beau.

*34 Tête imberbe janiforme, coiffée de la stéphané. ℞. Jupiter tenant un sceptre et lançant le foudre, à demi-nu, dans un quadrige au galop à d., conduit par la Victoire. B. I, p. 23, 26. El. Mm. 15, gr. 2,79. Superbe.

— Vente Mann, Londres 1917. —

*35 Tête imberbe d'Hercule à d., coiffée de la peau de lion; au-dessous, la massue. ℞. ROMA· Pégase au galop à d. Derrière lui, une massue. B. I, p. 29, 41. Mm. 20, gr. 7,40. Patine et conservation magnifiques.

36 Tête de Minerve à d., coiffée du casque phrygien se terminant en tête d'aigle. ℞. ROMA· Chien marchant à d. B. I, p. 29, 42. Sa. 1153. Æ. Mm. 7, gr. 1,21. Superbe pat. vert-bleu. Très beau.

37 Un deuxième exemplaire. Æ. Mm. 7, gr. 1,86. Beau.

Monnaies sans marque monétaire

MONNAIES DE LA REPUBLIQUE ROMAINE

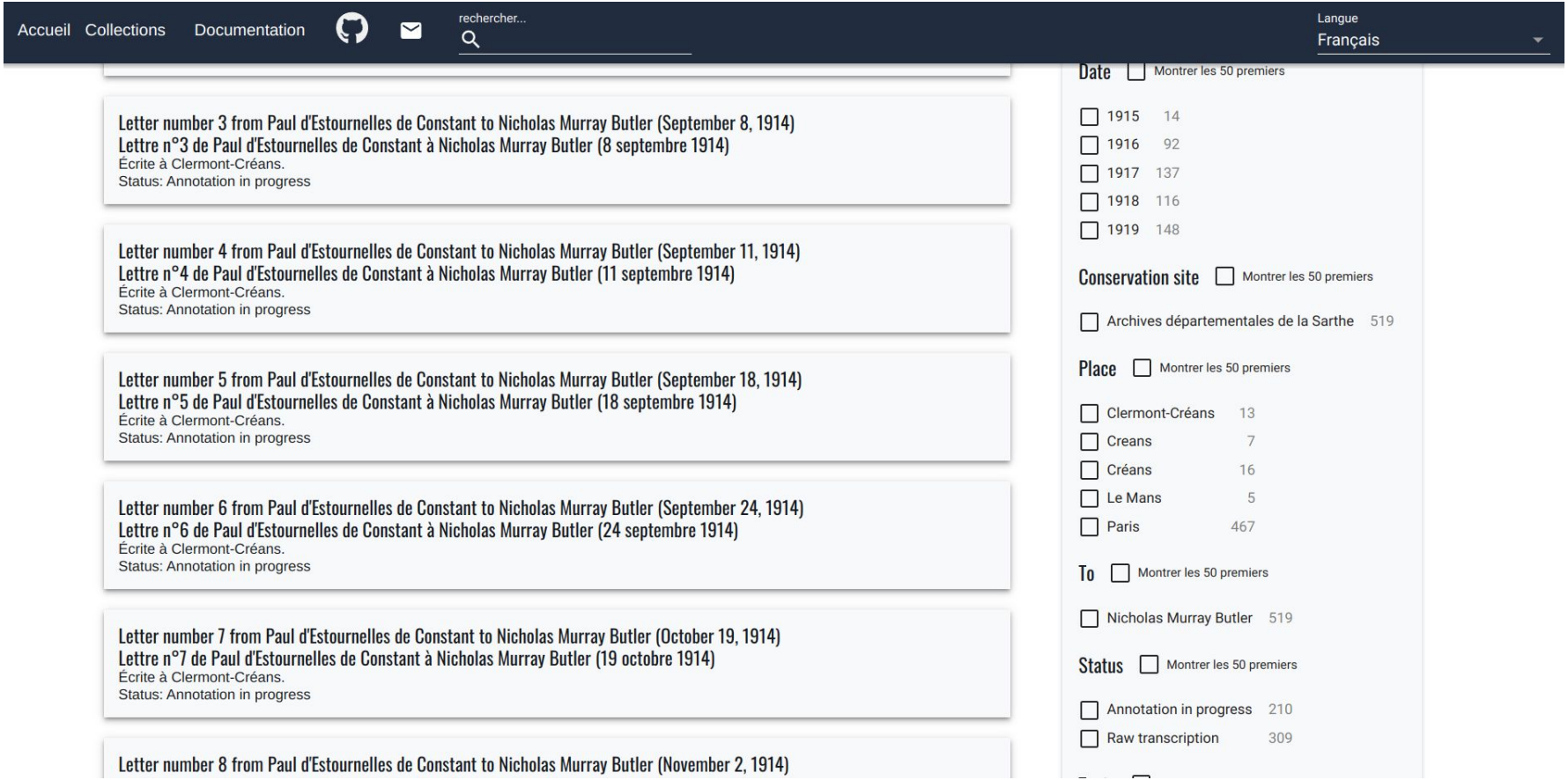
(B. = Babelon, Ernest, Monnaies de la République Romaine. 2 vol. Paris 1885.) - Toutes les pièces, sauf indication contraire, sont des deniers d'argent. -

Monnaies romano-campaniennes

342-211 av. J.-C.

- *33 La louve tournée à d., allaitant les jumeaux. A l'ex., -. ROMA. Corbeau à d., tenant une fleur dans son bec; derrière, B. I, p. 20, 20. Sambon, Italie, 1158. AE. Mm. 30, gr. 21,36. Pat. Très beau.
- *34 Tête imberbe janiforme, coiffée de la stéphané. -. Jupiter tenant un sceptre et lançant le foudre, à demi-nu, dans un quadrige au galop à d., conduit par la Victoire. B.

Exemple de visualisation basique avec TEI Publisher d'un catalogue de vente encodé en TEI



Correspondance de d'Estournelles de Constant, Floriane Chiffoleau (Inria, Le Mans Université), DiScholEd
https://dischole d.huma-num.fr/exist/apps/dischole d/index.html?collection=pec

Accueil Collections Documentation Télécharger rechercher... Langue Français

PAGE PRÉCÉDENTE TABLE DES MATIÈRES MÉTADONNÉES AFFICHAGE CITATION LICENCE PAGE SUIVANTE

Letter number 2 from Paul d'Estournelles de Constant to Nicholas Murray Butler (September 3, 1914)

Lettre suivante >

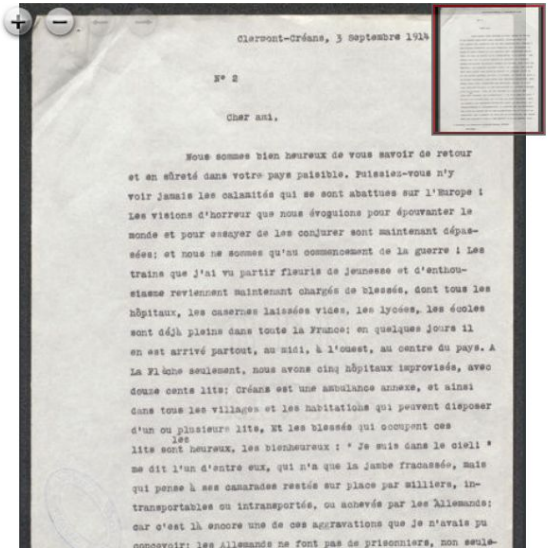
View
Reading version with critical

Clermont-Créans, 3 Septembre 1914

N° 2

Cher ami,

Nous sommes bien heureux de vous savoir de retour et en sureté dans votre pays paisible. Puissiez-vous n'y voir jamais les calamités qui se sont abattues sur l'Europe ! Les visions d'horreur que nous évoquions pour épouvanter le monde et pour essayer de les conjurer sont maintenant dépas-sées ; et nous ne sommes qu'au commencement de la guerre ! Les trains que j'ai vu partir fleuris de jeunesse et d'enthou-siasme reviennent maintenant chargés de blessés, dont tous les hôpitaux, les casernes laissées vides, les lycées, les écoles sont déjà pleins dans toute la France ; en quelques jours il en est arrivé partout, au midi, à l'ouest, au centre du pays. A La Flèche seulement, nous avons cinq hôpitaux improvisés, avec douze cents lits : Créans est une ambulance annexe, et ainsi dans tous les villages et les habitations qui peuvent disposer d'un ou plusieurs lits. Et les blessés qui occupent ces lits sont les heureux, les bienheureux : " Je suis dans le ciel! " me dit l'un d'entre eux, qui n'a que la jambe fracassée, mais qui pense à


Clermont-Créans, 3 Septembre 1914

N° 2

Cher ami,

Nous sommes bien heureux de vous savoir de retour et en sûreté dans votre pays paisible. Puissiez-vous n'y voir jamais les calamités qui se sont abattues sur l'Europe ! Les visions d'horreur que nous évoquions pour épouvanter le monde et pour essayer de les conjurer sont maintenant dépas-sées; et nous ne sommes qu'au commencement de la guerre ! Les trains que j'ai vu partir fleuris de jeunesse et d'enthou-siasme reviennent maintenant chargés de blessés, dont tous les hôpitaux, les casernes laissées vides, les lycées, les écoles sont déjà pleins dans toute la France; en quelques jours il en est arrivé partout, au midi, à l'ouest, au centre du pays. A La Flèche seulement, nous avons cinq hôpitaux improvisés, avec douze cents lits: Créans est une ambulance annexe, et ainsi dans tous les villages et les habitations qui peuvent disposer d'un ou plusieurs lits, Et les blessés qui occupent ces lits sont les heureux, les bienheureux : " Je suis dans le ciel! " me dit l'un d'entre eux, qui n'a que la jambe fracassée, mais qui pense à les camarades restés sur place par milliers, in-transportables ou intransportés, ou achevés par les Allemands; car c'est là encore une de ces aggravations que je n'avais pu

View
Index

ADD

Persons

Butler, Nicholas Murray
Born 02 April 1862 in Elizabeth (New Jersey)
Died 07 December 1947 in New York City (New York)
Nationality: American
Gender: Male
Occupation: philosopher/diplomat/educator
Education:
- Columbia University

Affiliation:
- President of the Columbia University (1902-1945)
- President of the Carnegie Endowment for International Peace (1925-1945)

Main event(s):
- Nobel Peace Prize (1931)

Correspondance de d'Estournelles de Constant, Floriane Chiffoleau (Inria, Le Mans Université), DiScholEd
https://discholed.huma-num.fr/exist/apps/discholed/pec/corpus/Lettre0005_18septembre1914.xml?root=2.4.2.2.7.6

Stage de fin d’études réalisé par Jules Nuguet (Université de Tours, M2 Intelligence des données de la culture et des patrimoines)

Publier les catalogues de vente avec TEI Publisher

- Cartographie du corpus des catalogues de vente,
- Modélisation en XML-TEI les métadonnées disponibles, à partir de la cartographie,
- Proposition d’une configuration de TEI Publisher répondant aux besoins de valorisation et d’accessibilité des données produites dans le cadre du projet. Développement d’une exploration de corpus avec des visualisation de données.

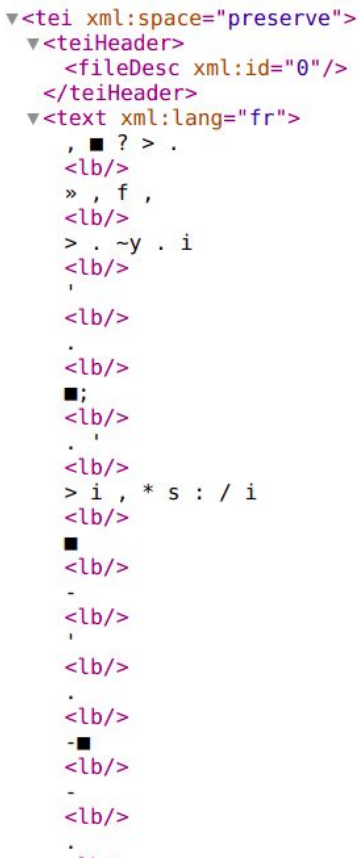
DataCatalogue et la science ouverte

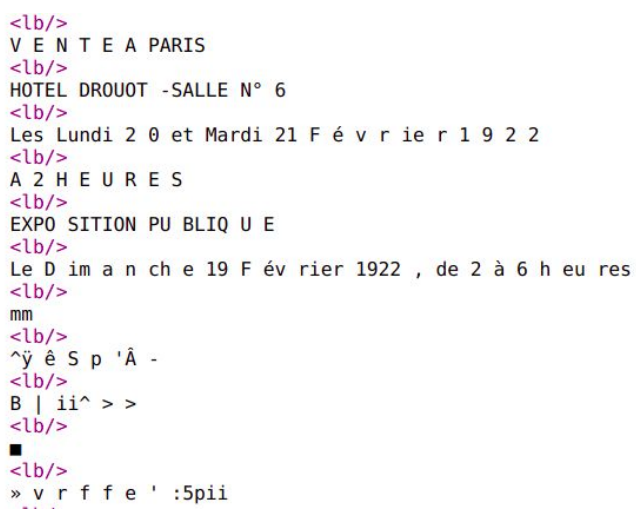
Tous les développements et les données créés dans le cadre du projet sont accessibles en ligne sur Github.

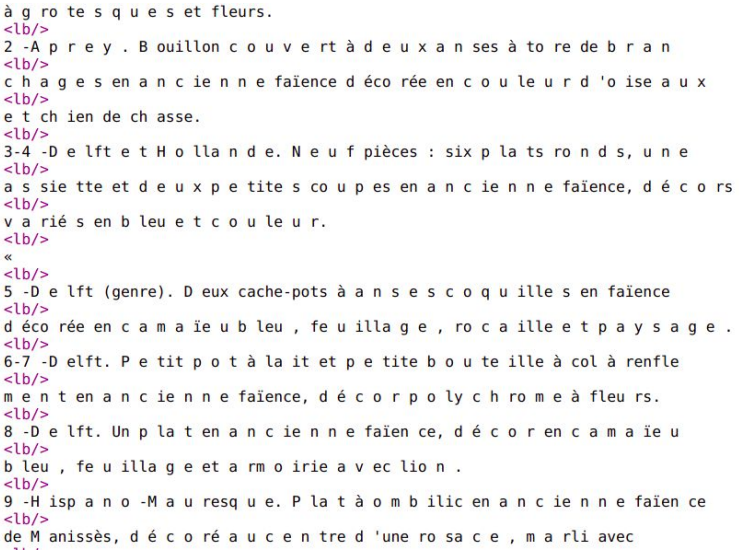
- Module GROBID DataCatalogue
- Sets d’entraînement
- Modélisation TEI

Les erreurs de transcription : un obstacle à la structuration automatique ?

Etude de la qualité du corpus en vue de la composition des sets d’entraînement

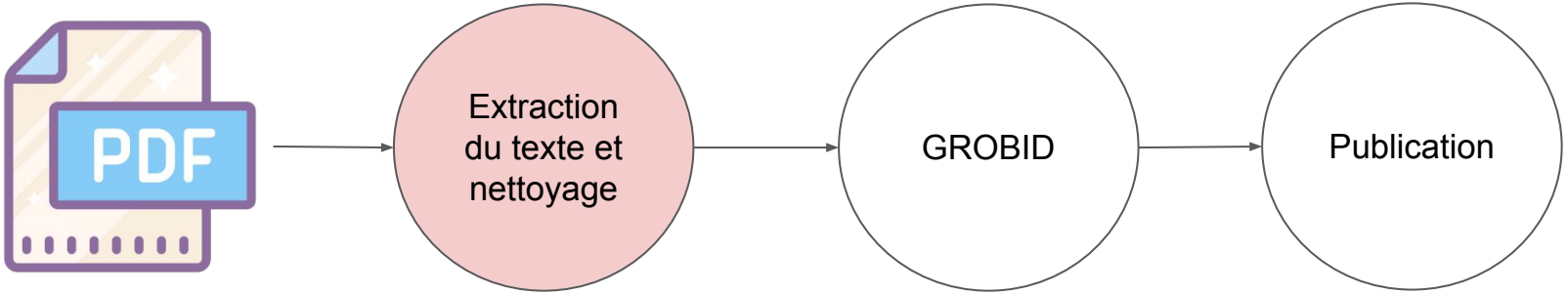


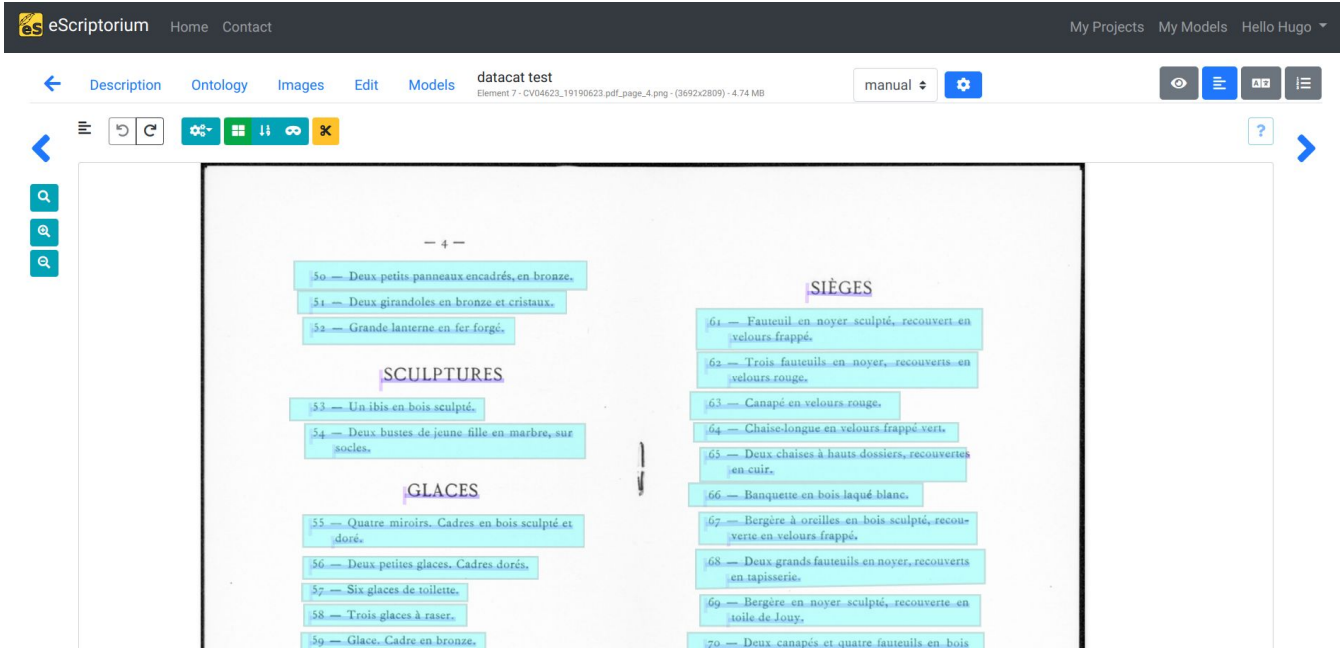




Exemple “extrême” des scories d’OCR rencontrés dans les PDF numérisés, ici de la collection Lair-Dubreuil. Ces erreurs ont des effets de cascades sur l’entraînement et l’inférence des modèles.

→ Envisager une étape de correction post-OCR des PDF pour nettoyer les scories générées durant la transcription automatique.





→ Segmentation de l'image avec eScriptorium et Kraken, donc non plus basée sur le texte mais sur la classification de pixel (le projet Gallic(orpor)a a notamment essayé cette méthode).

→ Segmentation de l'image avec de la détection d'objet, avec des outils type YOLOv5 (Clérice, 2022).

En conclusion, cette phase expérimentale de DataCatalogue a permis de :

- Créer un schéma TEI adapté à la représentation des catalogues de vente, qui permet de structurer finement l'information.
- Lancer le développement du module GROBID DataCatalogue, et d'obtenir des performances satisfaisantes pour les premiers modèles de segmentation. Les scores laissent une marge d'amélioration qui sera atteinte en raffinant et augmentant les exemples annotés.
- Créer des corpus annotés pour chaque niveau de segmentation développé,
- Débuter le prototypage d'une application de publication et de réfléchir au modèle de publication des fichiers restructurés.

Merci pour votre attention !

Contact : hugo.scheithauer[at]inria.fr