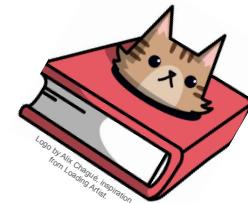


DataCatalogue : rétro-structuration automatique des catalogues de vente

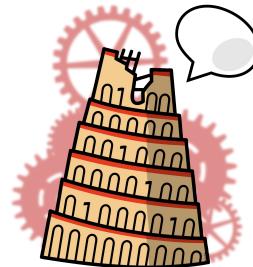
Hugo Scheithauer¹, Sarah Bénière¹, Jean-Philippe Moreux², Laurent Romary¹

¹ ALMAaCH - Automatic Language Modelling and ANAlysis & Computational Humanities, Inria Paris

² Bibliothèque Nationale de France



Logo By Alia Chapelle, Inspiration from LogoingArtist.



Institut
national
d'histoire
de l'art

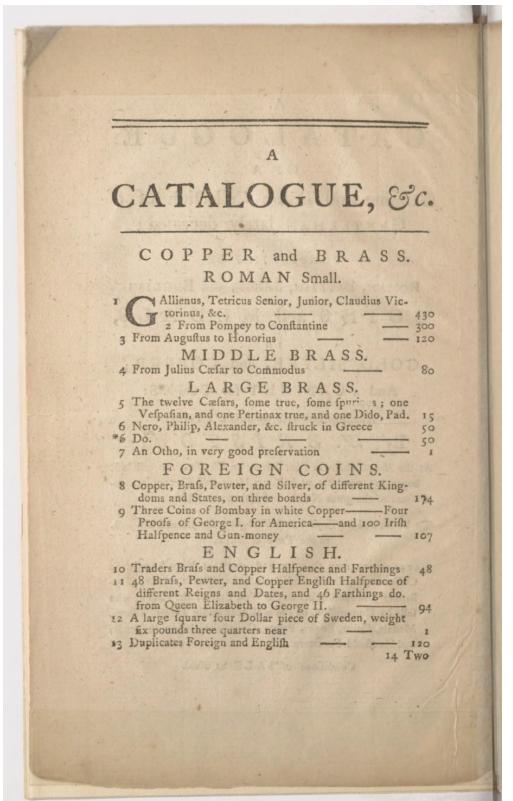


DataCatalogue : objectifs du projet

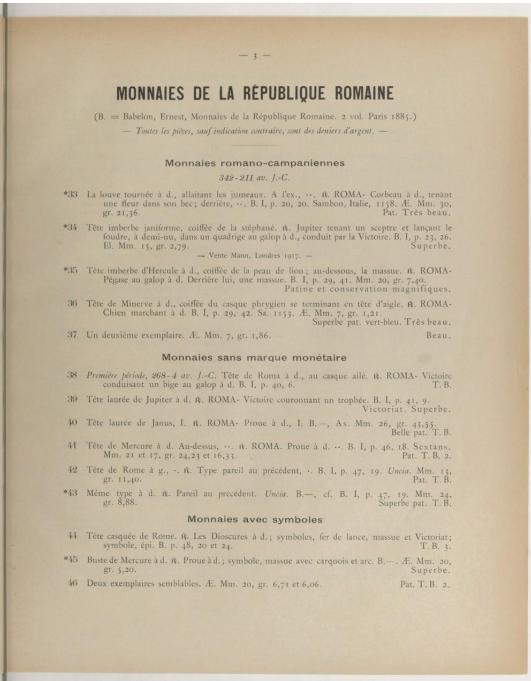
- Passer d'une **numérisation** à une **base de données textuelle et requêtable**.
- **Segmenter** les catalogues de vente et attribuer à chaque niveau d'information une **étiquette** : entrées de catalogues, numéro des entrées, description des objets, matériaux, sommes monétaires, etc.
- Produire, à 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.



Les catalogues de vente : homogénéité de mise en page



Whiston Bristow, 1762.



Lucien Naville, 1924.

Source gallica.bnf.fr / Bibliothèque nationale de France

60*
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. 1081). **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. (C. 1084). Or: Ens. 2 p. Très Beau.
700 / 800 €

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

50
LOUIS XVII (1791-1824). première restauration. **20 francs** à tête habillée.
Paris. 1814. (C. 1026). Seconde restauration. **20 francs**. Paris. 1817. (C. 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**. Ille. 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 litrai**. Tête féminine couronnée
à l'épis. R/ Justice marchant à g. (Coll. Jameson 860). Or. 18 g. Très Beau.
Rare.
Ancienne collection Feuardent.
1 200 / 1 500 €

57*
SICILE. **Drachme d'or**. Tête de Perséphone à g. R/ Victoire
allie conduisant un bœuf à g. (Coll. de Hirsch, 677. Cat. Gubkisen 345
var.). 4,24 g. Légères traces de monture sinon Très Beau.
Ancienne collection Feuardent.
3 200 / 3 600 €

58*
Béotie. Thibes. **Statère** (426-395). Boudier Iakobien. R/ Tête de Dionysos
barbu à dr. couronné de lierre. (PB 3523, SNC Del. 1356). Arg. 12,34 g.
Légèrement excentré sinon Très Beau.
Ancienne collection Feuardent.
300 / 400 €

Monnaies romaines

59*
JULES CESAR. **Aureus**. (46 avc.). Tête volée de Piatas à dr. R/ Litus, jarre et
hache. (Syd. 1017, BD 414). Or. 7,94 g. Très Beau à Supérieure.
2 000 / 2 500 €

60*
OTHON (15 janv. - mi avril 69). **Aureus**. Tête nue à dr. 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 volé. TB.
3 000 / 3 500 €

61*
Auguste. 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. 7,13 g. Traces de
restauration. Camée en flan sinon Très Beau. Joli style. Très rare.
5 500 / 6 000 €

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

63*
CONSTANCE II. **Siliqua réduite**. Lyon. (360-363). R/ VICTORIA DO NN
AVG. Victoire tenant une couronne et une palme. (RIC 210). Arg. Très Beau
à Supérieure.
200 / 250 €

Trésor de Morthomiers dans le Cher

La première partie de cette trouvaille d'antoniens effectuée au printemps
1953, fut déjà vendue le 23 septembre 1950, par l'éditeur Fraysse.

64
Antoniens. Caracalla (217-217). (C. 287). Elagabal (218-222). (C. 255).
Gordien III (4) (C. 404, C. 81, C. 86...). Philippe père (4) (C. 6, C. 17-1).
Valérien père (C. 6). Gallien (260-268) (C. 308). Arg. Ens. 12 p. TB et Très Beau.
80 / 150 €

65
Antoniens. Gordien III (238-244) (4). (C. 353, C. 404, C. 173...). Philippe
père (4) (C. 404, C. 81, C. 86...). Philippe père (2) (C. 216, C. 63).
Trajan (2) (C. 132). Valérien père (253-260) (2). (C. 25, C. 65). Arg. Ens. 12 p. TB et
Très Beau.
80 / 150 €

66
Antoniens. Gordien III (4) (C. 357, C. 404...). Philippe père. Philippe
père (4) (C. 404, C. 81, C. 86...). Valérien père (2) (C. 216, C. 63).
Brutus (2) (C. 19). Maniniane. Salomine (2) (C. 130...). Arg. Ens. 16 p. TB et
Très Beau.
150 / 200 €

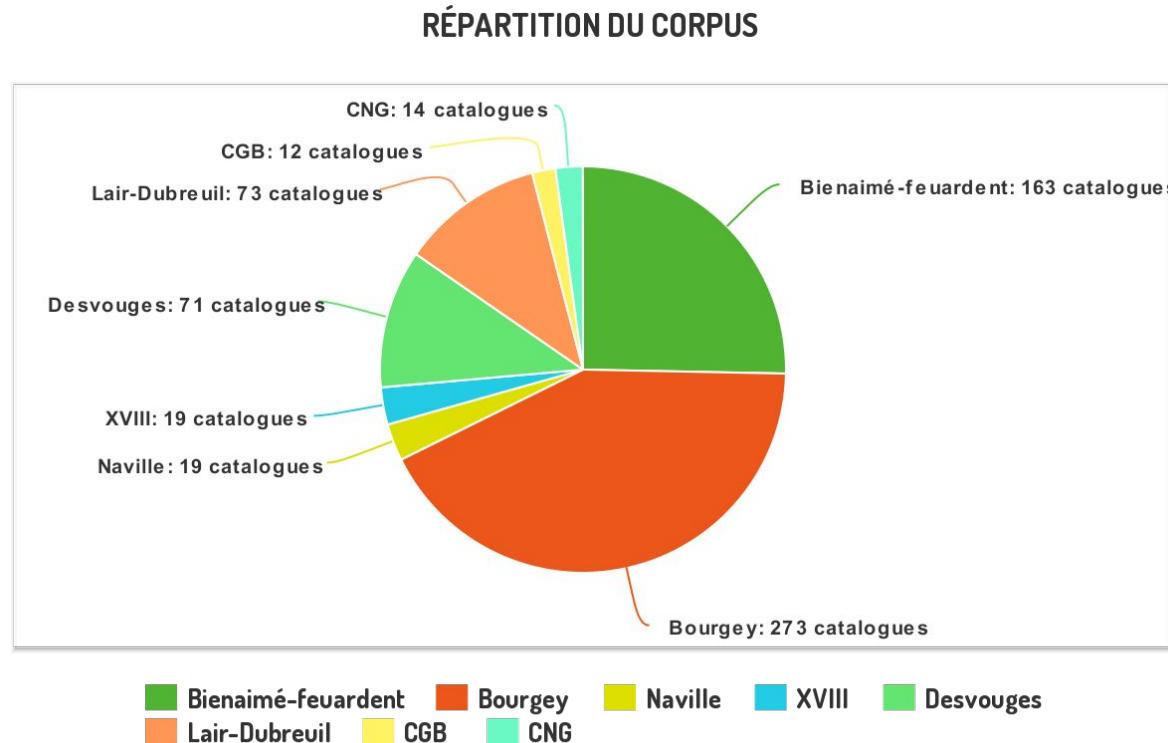
Fraysse & Associés, 2011.

Un premier échantillonnage des catalogues de vente de la BnF et de l'INHA : 4 siècles de contenus structurés

Siècles représentés : XVIIIe,
XIXe, XXe, XXIe.

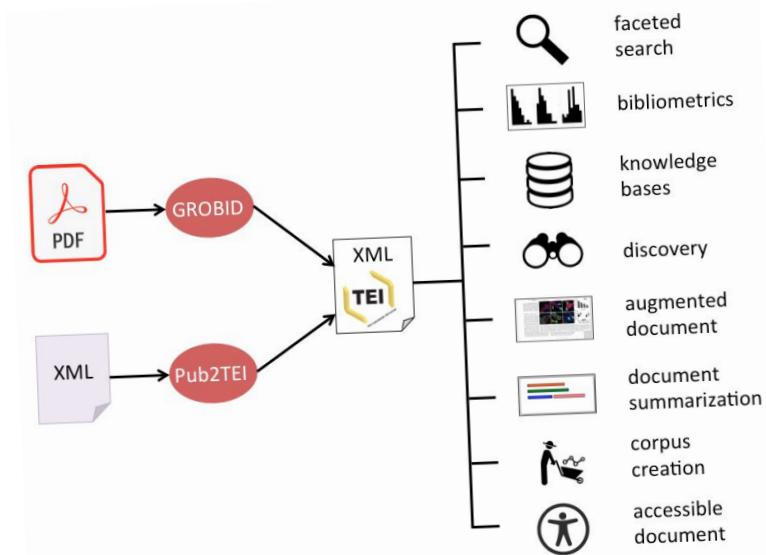
Langues : ~95% en français.

Types de vente :
numismatique, livres,
antiquités, objets d'art, objets
de luxe

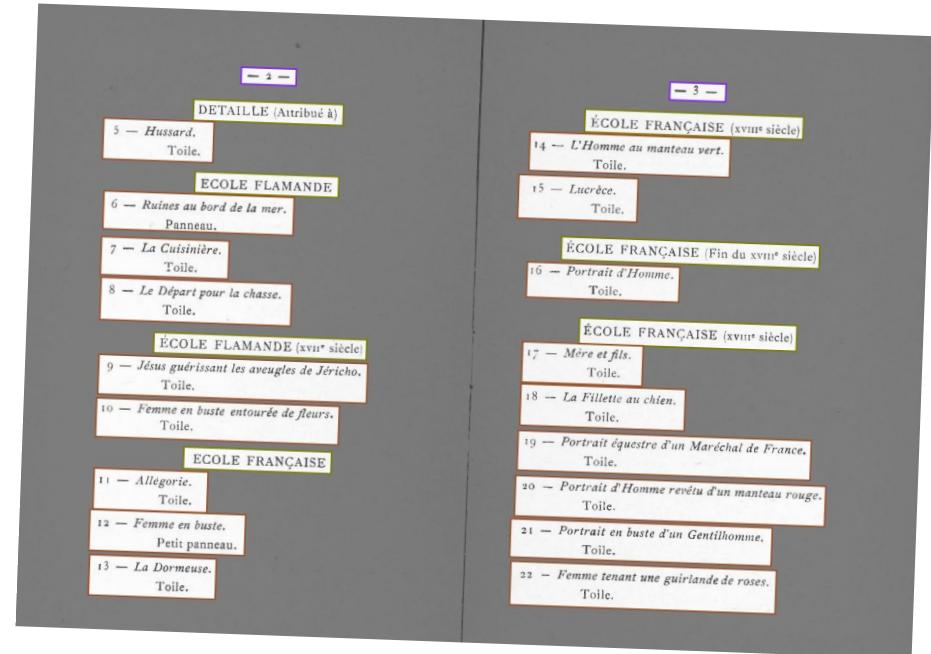


Quelles technologies pour gérer l'hétérogénéité de contenu et la diachronie des catalogues de vente ?

GROBID (GeneRation of Blibliographic Data)



Détection automatique d'objet



Quelles technologies pour gérer l'hétérogénéité de contenu et la diachronie des catalogues de vente ?

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

numéro d'items



description de l'objet mis en vente

*35

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

informations bibliographiques

B. I, p. 29, 41.

quantités et mesures

Mm. 20, gr. 7,40.

observations de conservation

Patine et conservation magnifiques.

Quelles technologies pour gérer l'hétérogénéité de contenu et la diachronie des catalogues de vente ?

147 — Grande miniature ovale : portrait d'homme à cheveux blancs avec manteau de fourrure; signée : *Eudelot*. Cadre doré.

Lair-Dubreuil, 1926 (INHA).

418 — F. 103. Peint. r. et bl. — Vulci. — Ext. *Les vaisseaux d'Ulysse passant devant les sirènes*. On voit deux vaisseaux voguant de front à pleines voiles. La proue est ornée d'une tête de sanglier, et à la poupe est une tête de cygne. Sur chaque vaisseau est représenté un homme debout et drapé, placé à la proue; un rameur est à l'arrière du navire; vers l'anse est une sirène placée sur un rocher, et retournant la tête vers les vaisseaux; plus loin est un dauphin. **NIKOSÖENES EΓΟΙΕ...** *Nicosithènes a fait.*

R. Le même sujet, si ce n'est que les proues sont ornées chacune d'un œil, et qu'aucun personnage ne paraît sur le devant des vaisseaux. Les rameurs sont dans la même position; seulement l'un d'eux lève la main comme surpris à l'aspect de la sirène placée sur un rocher, comme sur l'autre face, et avec un poisson plus loin.

- Les voiles des vaisseaux sont peintes en blanc.

Diam. 10 pouces 2 lignes. Haut. 4 pouces 8 lignes. 1795. 100 francs.

1	English, Irish, Manx, American, Swedish copper and leaden casts	45 0. 3. 0
2	Ditto	65 0. 6. 6
3	Ancient seals and casts	22 1. 11. 6
4	Four drawers, middle roman brases	140 0. 14. 0
5	Ditto, large and small, from Cæsar to Constantine	140 1. 14. 0

Langford, 1772 (BnF).

86 — **Drachme.** Mêmes types variés. - (166-88). **Drachme.** Tête radiée à dr. R. / Rose. - **CARIE.** HALICARNASSE. (1^{er} siècle av. J.-C.). **Obole.** Tête d'Hélios de face. R. / Tête d'Athéna à dr. (BMC. 153 sq. var., 247; P. 2610 var.). 2g71, 2g74, 0g89. Ens. 3 p. T.B. et Très Beau. 1.000/1.200

87 — **ILES DE CARIE. RHODES.** (166-88). **Bronze.** Tête radiée d'Hélios à dr. R. / Rose avec 2 branches. (BMC. 324 sq. - SNG. von Aulock 2835). Br. lg38. Patine brune. Très Beau. 1.000/1.500

Rollin, 1836 (BnF).

Bourgey, 1995 (BnF).

Standardiser la structuration des catalogues de vente



- Standard XML pour l'**édition de texte**, permet de rendre **lisible et compréhensible** un texte par un ordinateur.
- **Structuration** des catalogues en XML TEI, basé sur des **éléments existants** et standardisés, ainsi que sur des **nouveaux éléments** créés pour modéliser précisément l'objet catalogue, notamment la **notice**. Une notice structurée doit rendre compte de la matérialité de l'objet vendu.
- Une **notice** de catalogue décrit un objet matériel mis à la vente selon un ensemble de normes appartenant au monde de la **vente aux enchères**, ainsi que de la **discipline** dont provient l'objet (numismatique, archéologie, etc.).
- Le module TEI DataCatalogue se base sur les travaux de modélisation du projet **ArtI@S** et du projet **Ledoux**, et cherche à **affiner la représentation des catalogues**.

- Simon Gabay, Barbara Topalov, Caroline Corbières, Lucie Rondeau Du Noyer, Béatrice Joyeux-Prunel, et al.. Automating ArtI@S – extracting data from exhibition catalogues. EADH 2021 - *Second International Conference of the European Association for Digital Humanities*, Sep 2021, Krasnoyarsk, Russia. ([hal-03331838](#))
- Emmanuel Chateau Dutier, Caroline Corbières. A broader <object> content model for art history. Next Gen TEI, 2021 - TEI Conference and Members' Meeting, Oct 2021, Virtual, United States. ([hal-03654979](#))

D'une image à une base de données XML

COLLECTION D'UN AMATEUR

1*

PHILIPPE IV le Bel (1285-1314). Denier d'or à la masse. 1^{ère} 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 €

```
<catalogueEntry>
  <catalogueDesc>
    <head>Collection d'un amateur</head>
  </catalogueDesc>
  <!-- ... -->
  <catalogueItem>
    <altIdentifier>
      <idno>2</idno>
    </altIdentifier>
    <metamark>x</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>
```

Publier les catalogues de vente structurés



- Plateforme de **publication de fichier TEI**
- *Open source*
- Permet d'**indexer** un corpus, de le **visualiser**, et de le rendre **requêteable**
- Interface entièrement **personnalisable**
- Expertise au sein de l'équipe ALMAnaCH, Inria



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., ... à. 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. à. Jupiter tenant un sceptre et lançant le foudre, à demi-nu, dans un quadrigle 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. à. 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. à. ROMA. Chien marchant à d. B. I, p. 29, 42. Sa. 1153. AE. Mm. 7, gr. 1,21. Superbe pat. vert-bleu. Très beau.

37 Un deuxième exemplaire. AE. 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 quadrigle au galop à d., conduit par la Victoire. B.

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

DataCatalogue - v1 : bilan

- Modèle de données TEI pour représenter les catalogues de vente
- Annotation et entraînement de modèles GROBID pour la macro-structuration des catalogues, et la segmentation des notices
- Premières réflexions sur un modèle d'édition pour les catalogues de vente avec TEI Publisher
- Identification des défis à relever pour passer à l'échelle avec GROBID

MONNAIES ROMAINES (en or sauf indication contraire)			
*104 — RÉPUBLIQUE. <i>L. Cassius Longinus. Denier.</i> Tête voilée de Vesta à g. R/. Citoyen votant à g. (Sydenham 935 -Babelon Cassia 10 - Bourgey-Dépeyrot 327). <i>Rome</i> , 52/50. Arg. 4g09. Superbe.	700/900		
*105 — JULES CÉSAR. (*44 av. J.-C.) <i>Aureus.</i> Tête voilée de la Piété âgée à dr. R/. Instruments de sacrifice. (Cohen 3 -S. 1018 - B. Hirtia 2, Julia 23 - B.D. 414). <i>Rome</i> , 46. 8g35. Flan légèrement incurv. T.B.	5.000/6.000		
*106 — OCTAVE. (42-27) <i>Aureus.</i> Tête nue à g. R/. Quadriga triomphal au pas à dr. (C. 77 - Roman Imperial Coinage 30 - Giard 80) <i>Rome</i> , 29. 7g94. Revers rare. Très Beau.	8.000/10.000		
*107 — EMPIRE. AUGUSTE. (27 av. -14 ap. J.-C.) <i>Aureus.</i> Tête laurée à dr. R/. Caius et Lucius deb. tenant chacun une haste et un bouclier. (C. 42 - RIC. 350 - G. 1648). <i>Lyon et at auxiliaires</i> , 2 av. -4 ap. J.-C. 7g88. T.B.	3.500/4.000		
*108 — TIBÈRE. (14-37) <i>Aureus.</i> Tête laurée à dr. R/. Livia ou la Justice assise à dr. (C. 15 - RIC. 3 - G. 15). <i>Lyon</i> , 30. 7g71. Très Beau.	8.000/9.000		
*109 — CLAUDE I^e. (41-54) <i>Aureus.</i> Tête laurée à dr. R/. Némésis marchant à dr. tenant un caducée; devant, serpent. (C. 57 - RIC. 30 - G. 49). <i>Lyon</i> , 46/47. 7g82. Sauf tranche inégale, beau portrait. Très Beau.	8.000/9.000		
*110 — <i>Aureus.</i> Même tête. R/. SPQR/P.P/OB CS dans une couronne de chêne. (C. 95 - RIC. 47 - G. 73). <i>Lyon</i> , 51/52. 7g79. Sauf léger manque de la couronne, Très Beau.	7.000/9.000		
*111 — NÉRON. (54-58) <i>Aureus.</i> Tête laurée à dr. R/. La Santé assise à g. tenant une patère. (C. 317 - RIC. 53 - G. 236). <i>Rome</i> , 66/67. 7g30. Très Beau.	8.000/9.000		
*112 — <i>Aureus.</i> Tête nue à dr. R/. La Valeur deb. à g., tenant une haste et un parazonium. (C. 225 - RIC. 30 - G. 42). <i>Lyon</i> , 61/62. 7g15. Presque Très Beau.	7.000/8.000		
*113 — <i>Aureus.</i> Même tête. R/. EX S C dans une couronne de chêne. (C. 206 - RIC. 19 - G. 18). <i>Lyon</i> , 56/57. 7g63. T.B.	7.000/8.000		
*114 — <i>Aureus.</i> Autre varié. (C. 208 - RIC. 21 - G. 20). <i>Lyon</i> , 57/58. 7g51. T.B.	5.000/7.000		
*115 — <i>Aureus.</i> Tête laurée à dr. R/. Nérone et Messaline (ou Poppée) deb. à g. (ou Auguste et Livia?). (C. 42 - RIC. 41 - G. 200). <i>Rome</i> , 64/66. 7g10. T.B.	4.000/5.000		
*116 — GALBA. (68-69) <i>Sesterce.</i> Buste lauré et drapé à dr. R/. La Liberté deb. à g. tenant un bonnet et un sceptre. (C. 130 - RIC. 35). <i>Rome ou Gaule</i> . Br. 27g31. Patine vert foncé. Beau portrait. Très bel exemplaire.	12.000/14.000		
*117 — VESPASIEN. (69-79) <i>Aureus.</i> Tête laurée à dr. R/. La Paix assise à g. tenant une branche et un caducée. (C. —RIC. 10). <i>Rome</i> , 69/71. 7g27. Presque Superbe. Variété rare.	13.000/15.000		
*118 — <i>Aureus.</i> Tête laurée à g. R/. Vache marchant à dr. (C. 116 - RIC 97). <i>Rome</i> , 76. 7g25. T.B.	4.000/5.000		
*119 — <i>Aureus.</i> Tête laurée à dr. R/. Temple rond et 3 statues. (C. 580 var. - RIC. 69B). <i>Rome</i> , 73. 7g19. Traces de monture. B/T.B.	3.000/4.000		
120 — <i>Aureus.</i> Tête laurée à dr. R/. Temple rond et 3 statues. (C. 580 var. - RIC. 69B). <i>Rome</i> , 73. 7g19. Traces de monture. B/T.B.			

Bourgey, 1992 (BnF).

Des défis à relever pour DataCatalogue v2

```
<tei xml:space="preserve">
  <teiHeader>
    <fileDesc xml:id="0"/>
  </teiHeader>
  <text xml:lang="fr">
    , ■ ? > .
    <lb/>
    » , f ,
    <lb/>
    > . -y . i
    <lb/>
    .
    <lb/>
    <lb/>
    ■;
    <lb/>
    .
    <lb/>
    > i , * s : / i
    <lb/>
    ■
    <lb/>
    .
    <lb/>
    <lb/>
    .
    <lb/>
    -■
    <lb/>
    -
    <lb/>
    .
    <lb/>
    <lb/>
    V E N T E A PARIS
    <lb/>
    HOTEL DROUOT -SALLE N° 6
    <lb/>
    Les Lundi 20 et Mardi 21 Février 1922
    <lb/>
    A 2 H E U R E S
    <lb/>
    EXPO SITION PU BLIQUE
    <lb/>
    Le Dimanche 19 Février 1922 , de 2 à 6 heures
    <lb/>
    mm
    <lb/>
    ^y é S p 'Â -
    <lb/>
    B | ii^ > >
    <lb/>
    ■
    <lb/>
    » v r f f e ' :5pii
```

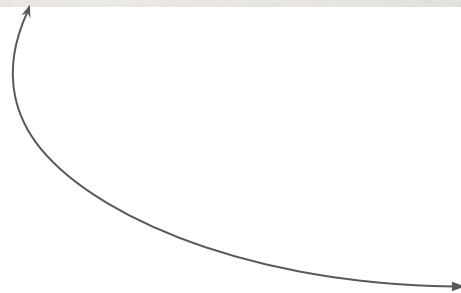
à g ro te s q u e s et fleurs.
<lb/>
2 -A p r e y . B ouillon c o u v e rt à d e ux a n ses à to re de b r a n
<lb/>
c h a g e s en a n c ie n n e faience d éco rée en c ou le u r d 'o ise aux
<lb/>
e t ch iem de ch asse.
<lb/>
3-4 -D e lft e t H olla n d e . N e u f piéces : six p la ts r o n d s , u n e
<lb/>
a s sie tte et d e ux p e tite s co up es en a n c ie n n e faience , d é c o rs
<lb/>
v a rié s en b leu e t c ou le u r .
<lb/>
``
<lb/>
5 -D e lft (genre) . D eux cache-pots à a n s e s c o q u ille s en faience
<lb/>
d éco rée en c a m a ie u b leu , fe u illa g e , ro c a ille e t p a y s a g e .
<lb/>
6-7 -D elft . P e tit p o t à la it et p e tite b ou te ille à col à renfle
<lb/>
m e n t en a n c ie n n e faience , d é c o r p o ly c h ro m e à fleu rs .
<lb/>
8 -D e lft . Un p la t en a n c ie n n e faien ce , d é c o r en c a m a ie u
<lb/>
b leu , fe u illa g e et a rm o irie a v ec lio n .
<lb/>
9 -H isp a n o -M a u r esq u e . P la t à o m b ilic en a n c ie n n e faien ce
<lb/>
de M anissé s , d é c o ré a u c e n tre d 'une ro sa c e , m a rli avec

Exemples “extrêmes” des scories d’OCR rencontrés dans les PDF numérisés : **caractères non existants dans les documents originaux, mots décomposées**. Ces erreurs ont des effets de cascades sur l’entraînement et l’inférence des modèles.

Des défis à relever pour DataCatalogue v2

*124 — **DOMITIEN.** (César, 69-81). **Aureus.** Tête laurée à dr. R/. La Santé nourrissant un serpent à dr. (C. 383 - RIC. 243). *Rome*, 79. 7g25. B/T.B. 3.500/4.000

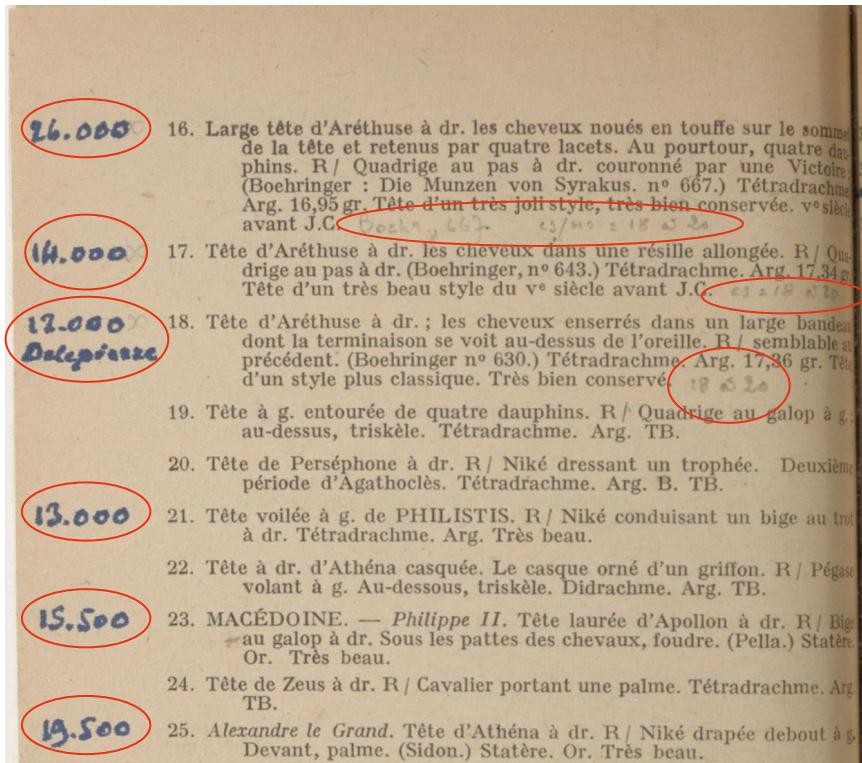
Bourgey, 1992, p. 28 (BnF).



→ Connecter les notices avec leurs illustrations respectives et en rendre compte dans l'encodage TEI

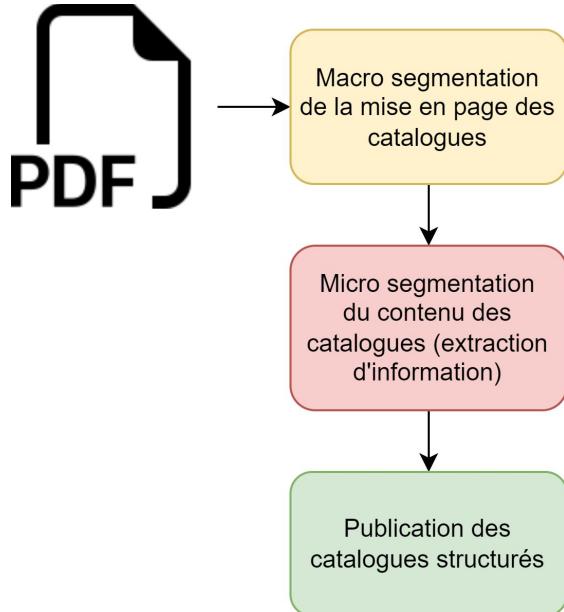
Bourgey, 1992, p. 29 (BnF).

Des défis à relever pour DataCatalogue v2



Segmentation, transcription et structuration des segments de texte manuscrits

DataCatalogue v2 - 2023-2024



{ BnF

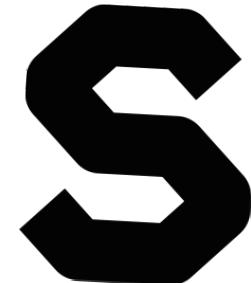


- Recrutement de Sarah Bénière dans l'équipe ALMAnaCH, Inria, en tant qu'ingénierie recherche et développement
- Crédit à Hugo Scheithauer pour la création d'un projet doctoral sur l'analyse automatique de mise en page mené par Paris PhD

Collaboration et réutilisation des données

- Campagne d'annotation de la macro-structure en coordination avec le projet **COLaF** (Inria, Multispeech)
- Schéma d'annotation suivant l'ontologie **SegmOnto**
- Corpus annotés et développements mis à disposition en ligne sur GitHub et Roboflow, selon les principes de la science ouverte
- Collaboration pluri-institutionnelles entre Inria, la BnF et l'INHA. L'idée est également de créer des corpus annotés et des outils facilement réutilisables par d'autres institutions

<https://github.com/DataCatalogue>



<https://segmonto.github.io/>



<https://colaf.huma-num.fr/>



<https://htr-united.github.io/>

Breaking news: #prixscienceouverte
#donneesrecherche Catégorie Prix jeunes chercheurs - 2023-11-29

Merci pour votre attention !

Ressources :

- GitHub DataCatalogue : <https://github.com/DataCatalogue>
- https://huggingface.co/spaces/HugoSchtr/DataCat_Yolov5
- Corpus de macro-segmentation des catalogues :
<https://app.roboflow.com/datacatalogue/macro-segmentation/overview>
- Guide d'annotation : <https://demo.hedgedoc.org/s/KRrbx3JP2>
- <https://datacat.hypotheses.org/category/segmentation>

Contacts :

sarah.beniere[at]inria.fr
hugo.scheithauer[at]inria.fr

