An ODD Schema for a Sustainable Encoding of Catalog Objects

Sarah Bénière, Hugo Scheithauer, Juliette Janes, Laurent Romary

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inria

Logo: Alix Chagué
Inspiration: Loading Artist

An ODD Schema for a Sustainable Encoding of Catalog Objects

Sarah Bénière, Research & Development Engineer, ALMAnaCH (Inria)

Hugo Scheithauer, PhD candidate, ALMAnaCH (Inria) & EPHE (Université PSL, Paris)

Juliette Janès, Research & Development Engineer, ALMAnaCH (Inria)

Laurent Romary, Research Director, DCIS (Inria)

The DataCatalogue Project

Institutional Context

Partnering institutions

- Inria
- (French) Ministry of Culture
- National Library (BnF)
- National Institute for Art History (INHA)

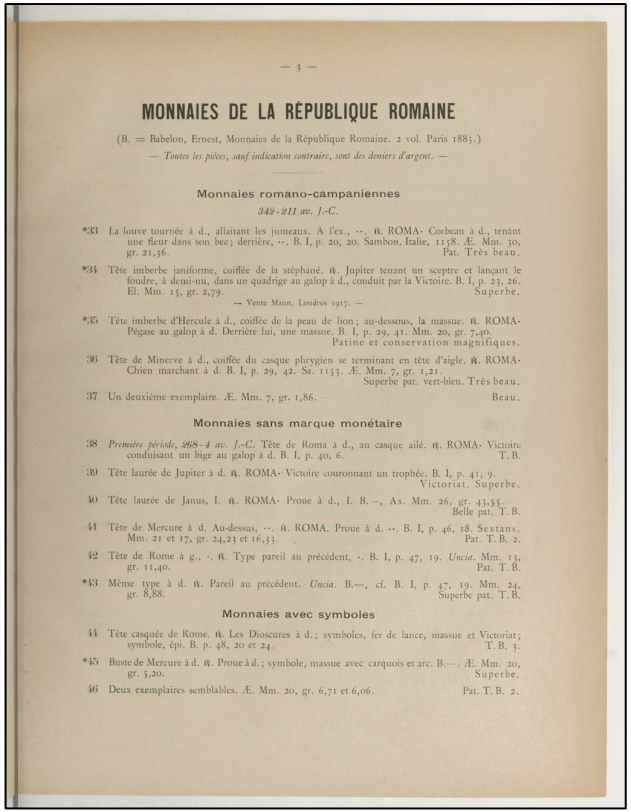
Project history

- Phase 1 (2021-2022)
- Phase 2 (2023-2024)

Scientific experts, researchers, data providers

Objective

Automate the **transformation** of **digitized sales catalogs** into a **structured database**, **using machine learning tools**



— 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

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.

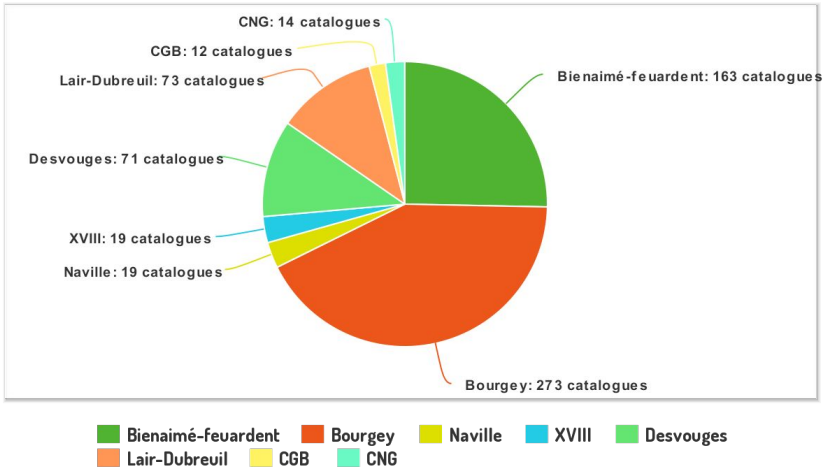
*45 Buste de Mercure à d. ℞. Proue à d.; symbole, massue avec carquois et arc. B.—. Æ. Mm. 20, gr. 5,20. Superbe.

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

Lucien Naville, 1924

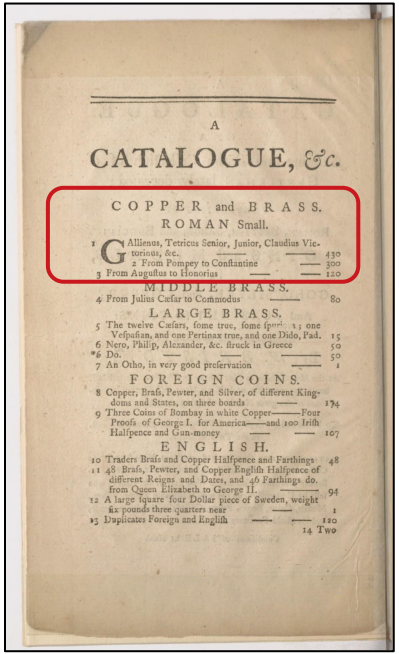
The Corpus

Composition



- Overall = +150k catalogs
- DataCatalogue dataset = **713 catalogs** (from BnF/INHA)
- Various representations of:
  - **Sales types** (numismatics, ancient books, antiques, fine arts, etc.)
  - **Document layouts**
  - **Time periods** (18th-21st centuries)
- Languages:
  - **French** (~95%)
  - German and English (~5%)

Interest – Heterogeneous Structure of Sales Catalogs



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 spurious; as one Vespasian, and one Pertinax true, and one Dido, Pad. 15
6 Nero, Philip, Alexander, &c. struck in Greece 50
*7 Do. —— —— 50
7 An Otho, in very good preservation —— 1

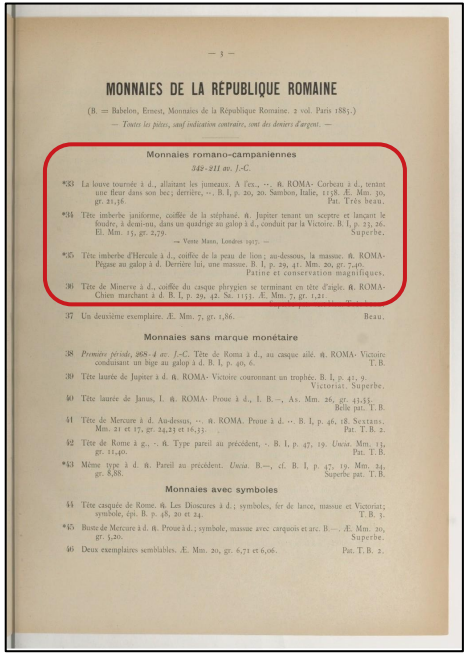
FOREIGN COINS.

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., ·-·. R. 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. R. 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. Pat. T. B. 2.
— Vente Mass, Londres 1917. —

*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. 7,49. Patine et conservation magnifiques.

36 Tête de Minerve à d., coiffée du casque phrygien se terminant en tête d'aigle. R. ROMA- Chien marchant à d. B. I, p. 29, 42. Sa. 1153. Æ. Mm. 7, gr. 1,21.

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

Monnaies sans marque monétaire

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

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

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

41 Tête de Mercure à d. Au-dessus, ·-·. R. 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., ·-·. R. Type pareil au précédent, ·-·. B. I, p. 47, 19. *Uncia*. Mm. 13, gr. 11,40. Pat. T. B.

*43 Même type à d. R. Pareil au précédent. *Uncia*. B.—, cf. B. I, p. 47, 19. Mm. 34, gr. 8,88. Superbe pat. T. B.

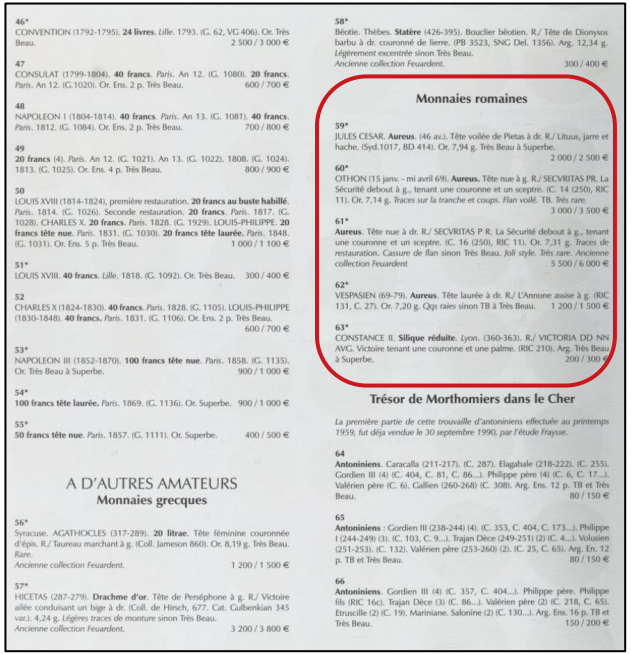
Monnaies avec symboles

44 Tête casquée de Rome. R. 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. R. Proue à d.; symbole, massue avec caducée 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 Frayssе.

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 €

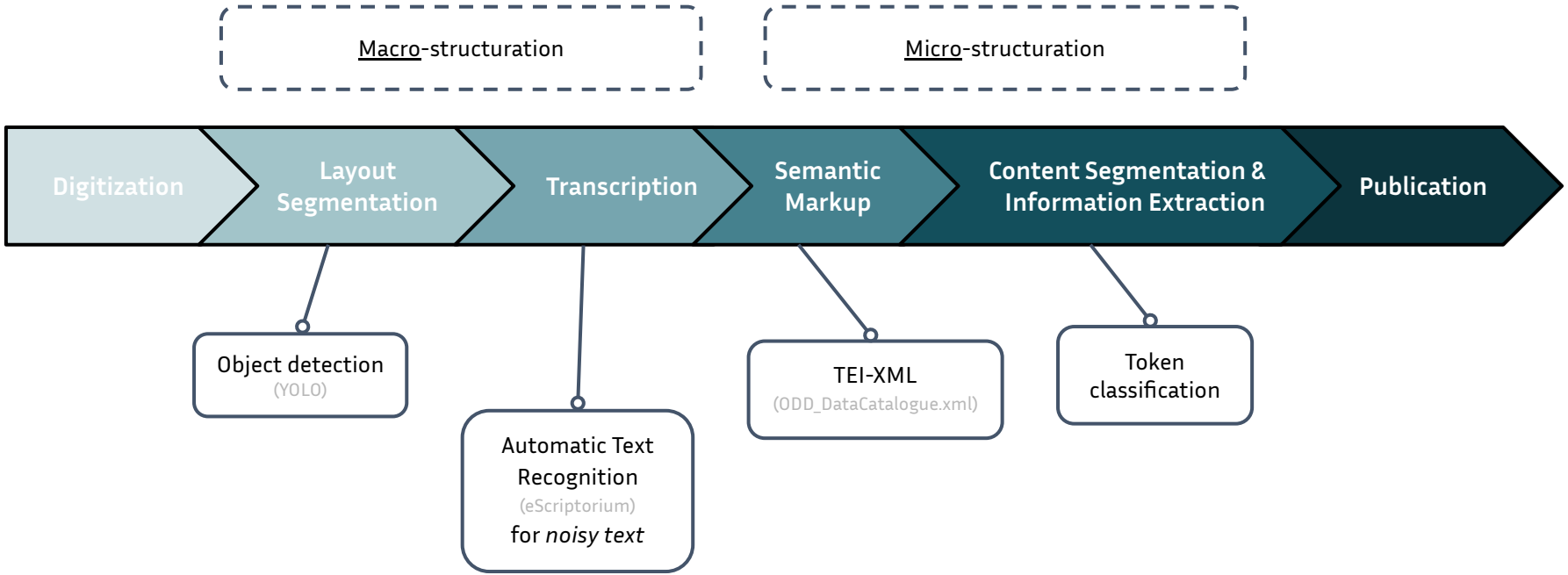
65
Antoniniens : Gordien III (238-244) (4). (C. 353, C. 404, C. 173...). Philippe I (244-249) (3). (C. 103, C. 9...). Trajan Dèce (249-251) (2) (C. 4...). Volusien (251-253). (C. 132). Valérien père (253-260) (2). (C. 25, C. 65). Arg. Ens. 12 p. TB et Très Beau. 80 / 150 €

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

The Workflow

The DataCatalogue Image-to-Text Workflow



Focus on Layout Segmentation

From SegmOnto to LADaS

SegmOnto

Gabay *et al.* (2023)

- Controlled vocabulary
- Layout description
- Specific syntax: **`Type:Subtype`**
- Documentation written in a TEI ODD

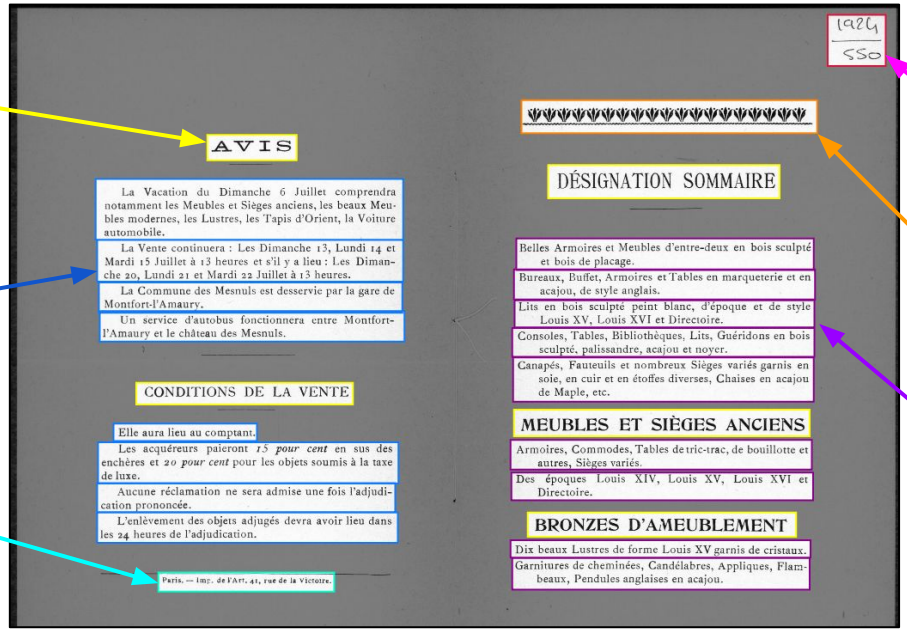
LADaS

Clérice *et al.* (2024)

Scheithauer *et al.* (2024)

- “Layout Analysis Dataset with SegmOnto”
- COLaF project (Inria)
- LADaS dataset = 5071 images
- DataCatalogue subset = **1425 images**

Example of DataCatalogue LADaS Classes



MainZone:Head

MainZone:P

MarginTextZone:Notes

MarginTextZone:ManuscriptAddendum

GraphicZone:Decoration

MainZone:Entry

Lair-Dubreuil, 1924

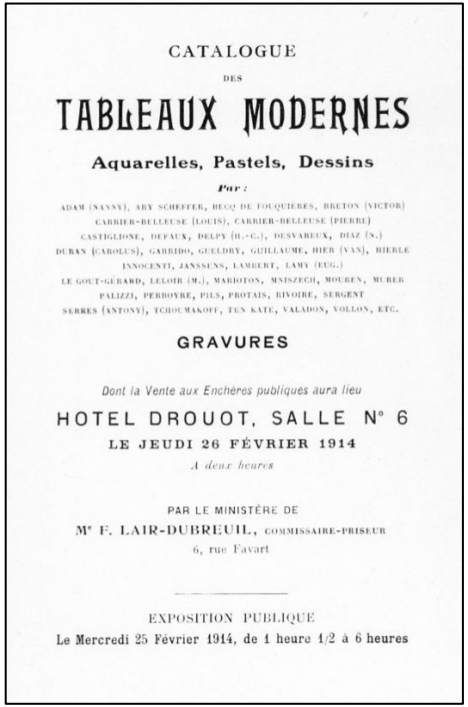
From LADaS to TEI

- **Table of equivalence** between LADaS and TEI-XML
- 2 “catalogue” classes:
  - `MainZone:Entry`
  - `MainZone:P@CatalogueDesc`
- 3 “catalogue” TEI elements:
  - `<catalogueItem>`
  - `<catalogueDesc>`
  - `<catalogueEntry>`

Class Name	TEI
GraphicZone	`<figure>`
GraphicZone:Decoration	`<figure type="decoration">`
GraphicZone:FigDesc	`<figDesc>`
MainZone:Entry	`<catalogueItem>`
MainZone:Head	`<head>`
MainZone:P	`<p>`
MainZone:P@CatalogueDesc	`<catalogueDesc>`
MainZone:Signature	`<signed>`
MarginTextZone:ManuscriptAddendum	`<div type="handwritten">`
MarginTextZone:Notes	`<note n="{}" place="{}">`
NumberingZone	`<pb n="" facs="">`
PageTitleZone	`<front>` → `<titlePage>`
	`<catalogueEntry>`

Describing Sales Catalogs in TEI

Title Page Encoded According to the P5 Guidelines



CATALOGUE

DES

TABLEAUX MODERNES

Aquarelles, Pastels, Dessins

Par :

ADAM (NANNY), ARY SCHEFFER, BECQ DE FOUQUIÈRES, BRETON (VICTOR) CARRIER-BELLEUSE (LOUIS), CARRIER-BELLEUSE (PIERRE) CASTIGLIONE, DEFAUX, DELPY (H.-C.), DESVAREUX, DIAZ (N.) DURAN (CAROLUS), GARRIDO, GUELDRY, GUILLAUME, HIER (VAN), HIERLE INNOCENTI, JANSSENS, LAMBERT, LAMY (EUG.) LE GOUT-GERARD, LELOIR (M.), MAIOTON, MNISZECH, MOUREN, MURER PALIZZI, PERBOYRE, PILS, PROTAIS, RIVOIRE, SERGENT SERRES (ANTONY), TCHOUMAKOFF, TEN KATE, VALADON, VOLLON, ETC.

GRAVURES

Dont la Vente aux Enchères publiques aura lieu

HOTEL DROUOT, SALLE N° 6

LE JEUDI 26 FÉVRIER 1914

A deux heures

PAR LE MINISTÈRE DE

Me F. LAIR-DUBREUIL, COMMISSAIRE-PRISEUR

6, rue Favart

EXPOSITION PUBLIQUE

Le Mercredi 25 Février 1914, de 1 heure 1/2 à 6 heures

Lair-Dubreuil, 1914

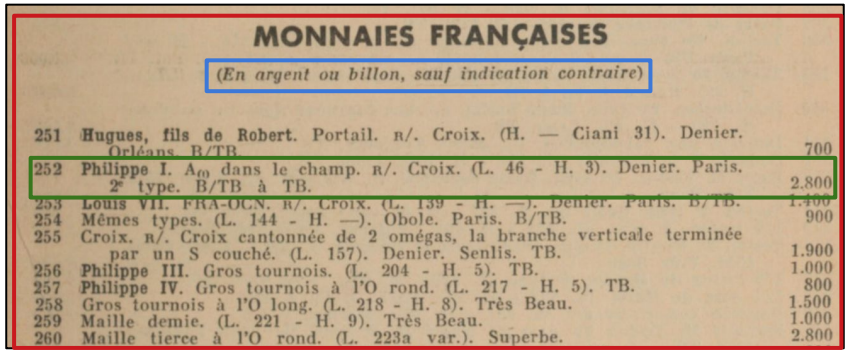
```
<front>
	<titlePage>
		<docTitle>
			<titlePart type="main">CATALOGUE <lb/>DES <lb/> TABLEAUX MODERNES</titlePart>
			<titlePart type="sub">Aquarelles, Pastels, Dessins</titlePart>
			<titlePart type="artists">Par : <lb/><name>ADAM (NANNY)</name>, <name>ARY SCHEFFER</name>,
				<name>BACQ DE FOUQUIERES</name>, <name>BRETON (VICTOR)</name>
				<lb/><name>CARRIER-BELLEUSE (LOUIS)</name>, <name>CARRIER-BELLEUSE (PIERRE)</name>
				<lb/><name>CASTIGLIONE</name>, <name>DEFAUX</name>, <name>DELPY (H.-C.)</name>,
				<name>DESVAREUX</name>, <name>DIAZ (N.)</name>
				<lb/><name>DURAN (CAROLUS)</name>, <name>GARRIDO</name>, <name>GUELDRY</name>,
				<name>GUILLAUME</name>, <name>HIER (VAN)</name>, <name>HIERLE</name>
				<lb/><name>INNOCENTI</name>, <name>JANSSENS</name>, <name>LAMBERT</name>, <name>LAMY
				(EUG.)</name>
				<lb/><name>LE GOUT-GÉRARD</name>, <name>LELOIR (M.)</name>, <name>MAIOTON</name>,
				<name>MNISZECH</name>, <name>MOUREN</name>, <name>MURER</name>
				<lb/><name>PALIZZI</name>, <name>PERBOYRE</name>, <name>PILS</name>,
				<name>PROTAIS</name>, <name>RIVOIRE</name>, <name>SERGENT</name>
				<lb/><name>SERRES (ANTONY)</name>, <name>TCHOUMAKOFF</name>, <name>TEN KATE</name>,
				<name>VALADON</name>, <name>VOLLON</name>, ETC.</titlePart>
			<titlePart type="sub">GRAVURES</titlePart>
			<titlePart type="sales_place">Dont la Vente aux Enchères publiques aura lieu
				<lb/><placeName>HOTEL DROUOT</placeName>, SALLE N°6 <lb/><date>LE JEUDI 26 FÉVRIER
				1914</date><lb/><time>A deux heures</time></titlePart>
		</docTitle>

		<byline>PAR LE MINISTÈRE DE <lb/><name>Me F. LAIR-DUBREUIL</name>, COMMISSAIRE-PRISEUR <lb/>
			<address>
				<addrLine>6, rue Favart</addrLine>
			</address>
		</byline>

		<metamark style="line"/>
		<titlePart><desc>EXPOSITION PUBLIQUE</desc>
			<lb/><date>Le Mercredi 25 Février 1914</date>, <time>de 1 heure 1/2 à 6 heures</time>
		</titlePart>
	</titlePage>
</front>
```

ExampleFile_Lair-Dubreuil_CV02553_19140226_f3.xml

TEI Modeling of Sales Catalogs

Building upon the grounding work of:

Gabay, Simon, Barbara Topalov, Caroline Corbières, Lucie Rondeau Du Noyer, Béatrice Joyeux-Prunel, and Laurent Romary. "Automating Artl@s – extracting data from exhibition catalogues." Presented at *EADH 2021 – Second International Conference of the European Association for Digital Humanities*, Krasnoyarsk, 2021. URL: https://hal.science/hal-03331838.



MONNAIES FRANÇAISES

(*En argent ou billon, sauf indication contraire*)

251 **Hugues, fils de Robert.** Portail. R/. Croix. (H. — Ciani 31). Denier. Orléans. B/TB. 700

252 **Philippe I.** A_{00} dans le champ. R/. Croix. (L. 46 - H. 3). Denier. Paris. 2e type. B/TB à TB. 2.800

253 **Louis VII.** FRA-OCN. R/. Croix. (L. 139 - H. —). Denier. Paris. B/TB. 1.400

254 Mêmes types. (L. 144 - H. —). Obole. Paris. B/TB. 900

255 Croix. R/. Croix cantonnée de 2 oméga, la branche verticale terminée par un S couché. (L. 157). Denier. Senlis. TB. 1.900

256 **Philippe III.** Gros tournois. (L. 204 - H. 5). TB. 1.000

257 **Philippe IV.** Gros tournois à l'O rond. (L. 217 - H. 5). TB. 800

258 Gros tournois à l'O long. (L. 218 - H. 8). Très Beau. 1.500

259 Maille demie. (L. 221 - H. 9). Très Beau. 1.000

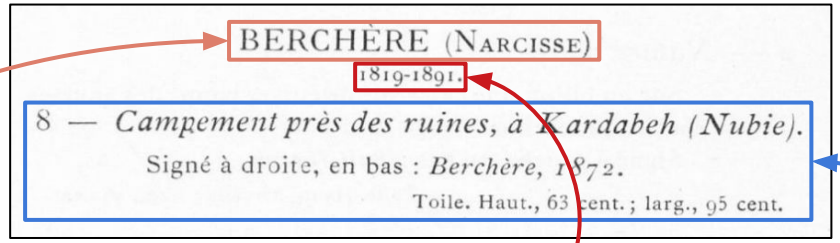
260 Maille tierce à l'O rond. (L. 223a var.). Superbe. 2.800

Bourgey, 1958

Description of Catalog-related Elements – <catalogueEntry>

- Closest content model ⇨ **model.divLike** (text division)
- Minimal structure:
  - **<head>** ⇨ header of the entry
  - **<catalogueDesc>** ⇨ further information, applies to all items in the entry [OPTIONAL]
  - **<catalogueItem>** ⇨ description of catalog objects (at least one)
- <catalogueEntry> elements can be embedded like elements

Example of an Encoded <catalogueEntry>



BERCHÈRE (NARCISSE)

1819-1891.

8 — *Campement près des ruines, à Kardabeh (Nubie).*

Signé à droite, en bas : *Berchère, 1872.*

Toile. Haut., 63 cent. ; larg., 95 cent.

Lair-Dubreuil, 1920

```
<catalogueEntry>
    <head>BERCHÈRE (Narcisse)</head>
    <catalogueDesc>1819-1891.</catalogueDesc>
    <catalogueItem n="8">
        <desc>
            <num>8</num> – Campement près de ruines, à Kardabeh (Nubie). <lb/>Signé à droite, en
            bas : Berchère, 1872. <lb/><objectType>Toile</objectType>. <dimensions><height>Haut.,
                63 cent.</height> ; <width>larg., 42 cent.</width></dimensions>
        </desc>
    </catalogueItem>
</catalogueEntry>
```

ExampleFile_Lair-Dubreuil_CV05182-19200318_f6.xmm

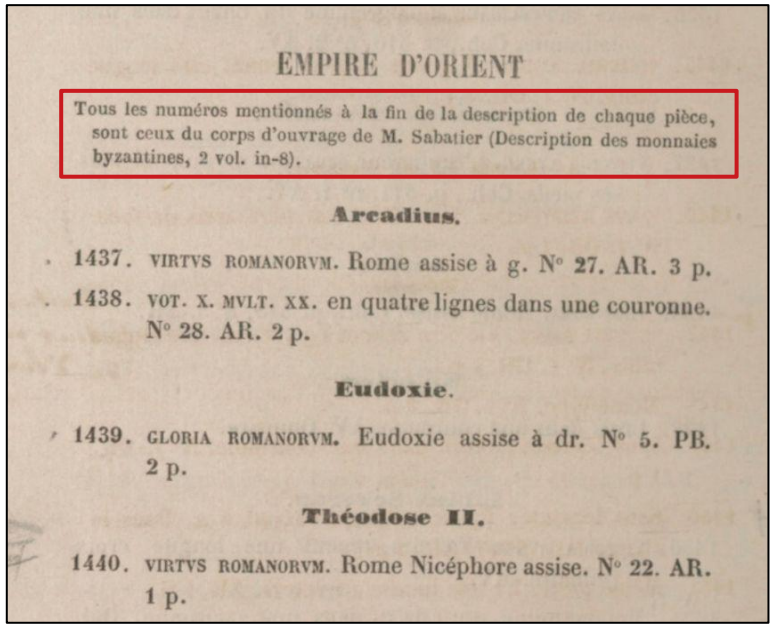
Embedding Challenge

- At which point do we switch from `<div>` to `<catalogueEntry>`?
- Do we consider “SANS NOM D’ATELIER” to be `<catalogueDesc>` or `<catalogueEntry>`?
- Does it depend on context?
- If so, can we actually automate the encoding?

```
<div>
        <head>METZ ET DOMAINES EPISCOPAUX</head>
        <div>
                <head>EVEQUES</head>
                <catalogueEntry>
                        <head>SEMI-EPISCOPALES - SEMI-IMPERIALES</head>
                        <catalogueEntry>
                                <head>Adalbéron I et Otton I, empereur (962-964)</head>
                                <catalogueDesc>SANS NOM D'ATELIER</catalogueDesc>
                        </catalogueEntry>
                        <catalogueEntry>
                                <head>Thiéri I et Otton I (964-973)</head>
                                <catalogueDesc> SANS NOM D'ATELIER</catalogueDesc>
                                <catalogueItem n="411"/>
                                <catalogueDesc>ATELIER DE METZ</catalogueDesc>
                                <catalogueItem n="412"/>
                        </catalogueEntry>
                </catalogueEntry>
        </div>
</div>
```

ExampleFile_bienaime-feuardent_12148-bpt6k9777420r_f119.xml

Description of Catalog-related Elements – `<catalogueDesc>`



EMPIRE D'ORIENT

Tous les numéros mentionnés à la fin de la description de chaque pièce, sont ceux du corps d'ouvrage de M. Sabatier (Description des monnaies byzantines, 2 vol. in-8).

Arcadius.

1437. VIRTVS ROMANORVM. Rome assise à g. N° 27. AR. 3 p.

1438. VOT. X. MVLT. XX. en quatre lignes dans une couronne. N° 28. AR. 2 p.

Eudoxie.

1439. GLORIA ROMANORVM. Eudoxie assise à dr. N° 5. PB. 2 p.

Théodose II.

1440. VIRTVS ROMANORVM. Rome Nicéphore assise. N° 22. AR. 1 p.

Rollin, 1864

- Closest content model ⇨ **`model.pLike`** (paragraph)
- LADaS equivalent ⇨ **`MainZone:P@CatalogueDesc`**
- Gives **further information** about the entry
- Content of **`<catalogueDesc>`** applies to all items in the entry

Description of Catalog-related Elements – <catalogueItem>

- Closest content model ⇨ **model.descLike** (description)
- LADaS equivalent ⇨ **MainZone:Entry**
- Information is structured within <desc>:
  - **<num>** ⇨ object identifier within the catalog
  - **<objectType>** ⇨ type of object described (e.g. painting, sculpture, coin, etc.)
  - **<dimensions>** ⇨ object dimensions
  - **<condition>** ⇨ assessment of the object's physical condition
- All elements within <desc> are **optional due to the heterogeneity of the entries' presentation**.

```
<catalogueItem n="8">
        <desc>
                <num>8</num> – Campement près de ruines, à Kardabeh (Nubie). <lb/>Signé à droite, en
                bas : Berchère, 1872. <lb/><objectType>Toile</objectType>. <dimensions><height>Haut.,
                                63 cent.</height> ; <width>larg., 42 cent.</width></dimensions>
        </desc>
</catalogueItem>
```

ExampleFile_Lair-Dubreuil_CV05182-19200318_f6.xmm

Conclusion & Perspectives

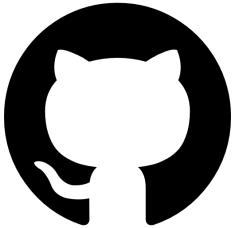
Key Takeaways

- The DataCatalogue workflow is still a work in progress:
  - Automate the transition from one step to another as much as possible
  - Resolve the embedding challenge

- Using the SegmOnto controlled vocabulary for layout segmentation ensures:
  - High compatibility with TEI
  - Standardized annotations and encoding
    - ➥ Improves DataCatalogue's commitment to make data FAIR

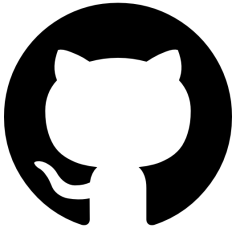
Useful Links

Read more about the DataCatalogue workflow

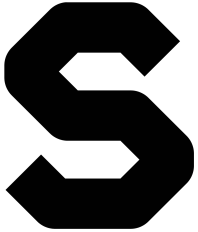


Check out the DataCatalogue annotation guide (FR)

Test the DataCatalogue object detection model



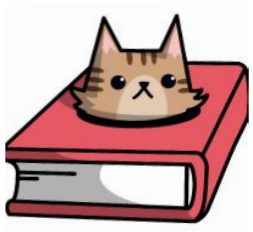
Access the DataCatalogue TEI repository



Read the SegmOnto documentation

Access the LADaS dataset

Thank you for your attention!



Logo: Alix Chagué
Inspiration: Loading Artist

Contact: sarah.beniere@inria.fr