
ALTO tags Use Cases

Version 16-01-2014



ALTO Board

Copyright © 2014 ALTO Board. All rights reserved.

No part of this publication may be reproduced, stored in databases, or transferred in any form (electronically, photo-mechanically, chemically, manually, or otherwise) without the express written permission of the ALTO Board. The software described in this manual is licensed software that may be used only in compliance with the licensing terms and conditions. The ALTO Board reserves the right to make changes to the content of this manual without notice. The ALTO Board makes no guarantee regarding the accuracy of the information provided in this manual. Microsoft, MS-DOS, and Windows are registered trademarks of the Microsoft Corporation.

Parts of the software uses the Duden-Proof-Factory of the Brockhaus Duden Neue Medien GmbH for syllable separation.

Product or company names that are mentioned may be trademarks or registered trademarks of the respective company. The ALTO Board uses these names and trademarks in the following manual merely for explanatory purposes and for the benefit of the respective user, and such use does not imply trademark infringement.

Under this software license, you are only permitted to reproduce materials that are not protected by copyright laws. This excludes only materials where you hold the copyright and/or legal permission to reproduce copyrighted materials. If you are uncertain about the copyright status of certain materials then please seek legal counsel. The ALTO Board holds no liability over copyright violations resulting from the use of this software.

Last updated: 15.01.2014

The ALTO Board



Table of Contents

1.	Introduction	4
1.1.	XML schema	4
1.2.	Good practices, limitations	4
1.3.	Uses cases	4
2.	Layout tagging	5
2.1.	XML schema	5
2.1.	Good practices, limitations	5
3.	Structural tagging	21
3.1.	XML schema	21
3.2.	Good practices, limitations	21
4.	Named Entities tagging	45
4.1.	XML schema	45
4.2.	Good practices, limitations	45
5.	Reference documents	49

1. Introduction

The ALTO version 2.1 introduces a new mechanism: tag (or annotation). The aim of this mechanism is to cover several change requests concerning the ALTO format:

- *Layout labelling* – ALTO format should record various region types (tables, graphics, maths formulas, music scores, etc.)
- *Logical Labelling Of Structural Elements* – Structural tagging of documents offers a large number of benefits. Full text search can be done in a much more focused way, reprinting of digitised books could be made easier, etc.
- *Named Entity markup* – Post-OCR processing on page images often includes named-entity recognition attempts.

■ XML schema

This mechanism has been introduced in ALTO version 2.1 schema. In XML terms, a reference to a new element group "Tags/Tag" allows to add additional information to the contents which are referring to these tag elements:

```
<Tags>
  <NamedEntityTag ID="NE15" LABEL="Location" DESCRIPTION="Lexington"/>
  ...
</Tags>

<Layout>
  ...
  <String CONTENT="Lexington" WC="1.0" TAGREFS="NE15" HPOS... VPOS...>
  ...
</Layout>
```

■ Good practices, limitations

- The tag reference must be set on the *upper* level available (eg on TextBlock, not on all TextLines or Strings of the block).
- A content crossing two or more pages can be described: the blocks composing this content must reference the same tag ID.
- If the IDs of the ALTO content elements change, the tag references must be updated.

■ Use cases

For readability purpose, IDs have been omitted in the following examples.

2. Layout tagging

Before the actual character recognition starts, any OCR software analyzes the page structure to identify regions. Modern OCR software supports various region types than ALTO could record: tables, graphics, music scores, etc. Some of these regions can be categorized for the OCR software to use a specific algorithm or process when recognizing their characteristics; they can also be identified with a manual labelling.

XML schema

ALTO tag element: <LayoutTag>

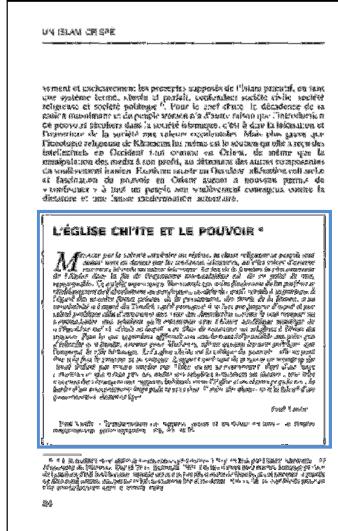
Tag attributes:

- **ID**: Tag id
- **LABEL**: Region type
- **TYPE** (optional): Supertype or category of the region type
- **DESCRIPTION** (optional): Description of the region, description of the region content, or the content itself
- **XmlData** (optional): Any XML encoded metadata helping to describe the tag, or the vocabulary used to describe the tag
- **URI** (optional): Any authority URI relevant for the tag

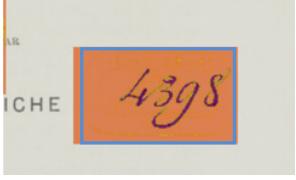
Good practices, limitations

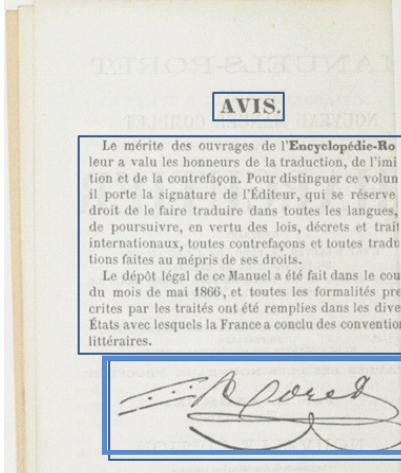
- The intent of layout tagging is to identify types of content. Sometimes, specific types of content can also have a functional value (sidebars, separators, etc.) and the question of using a layout tag or a structural tag may be raised.

Text tags, to be applied on <ComposedBlock>, <TextBlock>, <TextLine> or <String> elements

Use cases	ALTO tag	Examples	Remarks
Frames, boxes, sidebars	<pre><LayoutTag LABEL="Sidebar" DESCRIPTION="Textual box"/></pre>	 <p>The screenshot shows a page from a document. At the top, there is a header section with the title "L'ISLAM EN SÉRIE". Below this, the main content begins with a section titled "L'ÉGLISE CHRÉTIENNE ET LE POUVOIR". This section contains several paragraphs of text. To the left of the main text block, there is a rectangular box with a thin black border. This box is labeled "L'ÉGLISE CHRÉTIENNE ET LE POUVOIR" at the top. Inside the box, there is text and a small blue arrow pointing towards the left edge of the box, indicating its position relative to the main text area.</p>	

Tables	<pre><LayoutTag LABEL="Table" DESCRIPTION="Table of numbers "/></pre>	<p>Situation de la Banque d'Angleterre, aux dates suivantes :</p> <table border="1"> <thead> <tr> <th></th> <th>18 juv. 49</th> <th>11 nov. 49</th> <th>4 mai 50</th> <th>15 juill. 50</th> <th>22 sept.</th> </tr> </thead> <tbody> <tr> <td>Total</td> <td>89,371,965</td> <td>77,899,210</td> <td>64,946,765</td> <td>53,218,020</td> <td>39,299,473</td> </tr> <tr> <td>Billets émis</td> <td>89,371,965</td> <td>77,899,210</td> <td>64,946,765</td> <td>53,218,020</td> <td>39,299,473</td> </tr> <tr> <td>Arct. Débits du gouvernement</td> <td>8,426,893</td> <td>8,426,893</td> <td>3,452,800</td> <td>3,452,800</td> <td>3,452,800</td> </tr> <tr> <td>Autres débits</td> <td>14,882,259</td> <td>13,749,259</td> <td>13,293,768</td> <td>13,118,010</td> <td>13,429,473</td> </tr> <tr> <td>Mouvements d'or et d'argent</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Totaux</td> <td>89,371,965</td> <td>77,899,210</td> <td>64,946,765</td> <td>53,218,020</td> <td>39,299,473</td> </tr> <tr> <td>Département des Finances</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Paior. Capital des administrateurs</td> <td>12,353,000</td> <td>14,838,000</td> <td>11,833,000</td> <td>11,833,000</td> <td>11,833,000</td> </tr> <tr> <td>Réserves</td> <td>3,000,000</td> <td>3,830,000</td> <td>3,830,000</td> <td>3,830,000</td> <td>3,830,000</td> </tr> <tr> <td>Autres... ...de la Banque</td> <td>14,736,117</td> <td>13,206,916</td> <td>13,195,000</td> <td>13,118,010</td> <td>13,429,473</td> </tr> <tr> <td>Autres dépôts</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>États à payer à 7 jours et moins</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Totaux</td> <td>16,089,117</td> <td>17,163,973</td> <td>17,152,900</td> <td>17,076,480</td> <td>17,114,786</td> </tr> <tr> <td>Arct. Généralités... ...de la Banque</td> <td>19,011,897</td> <td>20,114,297</td> <td>17,450,907</td> <td>19,411,486</td> <td>19,729,433</td> </tr> <tr> <td>Billets de Banque en poche/bourse</td> <td>8,426,893</td> <td>8,426,893</td> <td>3,452,800</td> <td>3,452,800</td> <td>3,452,800</td> </tr> <tr> <td>Mouvements d'or et d'argent</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Totaux</td> <td>41,006,817</td> <td>43,681,207</td> <td>39,439,377</td> <td>39,446,951</td> <td>39,299,473</td> </tr> </tbody> </table> <p>Circulation des notes des Banques du Roy.-Uni, aux dates suivantes :</p> <table border="1"> <thead> <tr> <th></th> <th>18 juv. 49</th> <th>11 nov. 49</th> <th>4 mai 50</th> <th>15 juill. 50</th> <th>22 sept.</th> </tr> </thead> <tbody> <tr> <td>Total</td> <td>37,077,375</td> <td>30,187,458</td> <td>25,769,510</td> <td>21,418,452</td> <td>19,956,740</td> </tr> <tr> <td>Banque d'Angleterre</td> <td>37,077,375</td> <td>30,187,458</td> <td>25,769,510</td> <td>21,418,452</td> <td>19,956,740</td> </tr> <tr> <td>Banques provinciales</td> <td>1,270,000</td> <td>1,270,000</td> <td>1,270,000</td> <td>1,270,000</td> <td>1,270,000</td> </tr> <tr> <td>Banques régionales (plus Bank)</td> <td>8,378,130</td> <td>8,366,587</td> <td>3,012,000</td> <td>3,819,070</td> <td>3,817,373</td> </tr> <tr> <td>... d'Écosse</td> <td>8,416,461</td> <td>8,399,207</td> <td>3,001,780</td> <td>3,815,079</td> <td>3,819,070</td> </tr> <tr> <td>... d'Irlande</td> <td>8,392,240</td> <td>8,366,617</td> <td>3,000,240</td> <td>3,814,000</td> <td>3,819,070</td> </tr> <tr> <td>Totaux</td> <td>37,077,375</td> <td>30,186,190</td> <td>26,361,382</td> <td>30,147,142</td> <td>30,360,185</td> </tr> </tbody> </table>		18 juv. 49	11 nov. 49	4 mai 50	15 juill. 50	22 sept.	Total	89,371,965	77,899,210	64,946,765	53,218,020	39,299,473	Billets émis	89,371,965	77,899,210	64,946,765	53,218,020	39,299,473	Arct. Débits du gouvernement	8,426,893	8,426,893	3,452,800	3,452,800	3,452,800	Autres débits	14,882,259	13,749,259	13,293,768	13,118,010	13,429,473	Mouvements d'or et d'argent						Totaux	89,371,965	77,899,210	64,946,765	53,218,020	39,299,473	Département des Finances						Paior. Capital des administrateurs	12,353,000	14,838,000	11,833,000	11,833,000	11,833,000	Réserves	3,000,000	3,830,000	3,830,000	3,830,000	3,830,000	Autres... ...de la Banque	14,736,117	13,206,916	13,195,000	13,118,010	13,429,473	Autres dépôts						États à payer à 7 jours et moins						Totaux	16,089,117	17,163,973	17,152,900	17,076,480	17,114,786	Arct. Généralités... ...de la Banque	19,011,897	20,114,297	17,450,907	19,411,486	19,729,433	Billets de Banque en poche/bourse	8,426,893	8,426,893	3,452,800	3,452,800	3,452,800	Mouvements d'or et d'argent						Totaux	41,006,817	43,681,207	39,439,377	39,446,951	39,299,473		18 juv. 49	11 nov. 49	4 mai 50	15 juill. 50	22 sept.	Total	37,077,375	30,187,458	25,769,510	21,418,452	19,956,740	Banque d'Angleterre	37,077,375	30,187,458	25,769,510	21,418,452	19,956,740	Banques provinciales	1,270,000	1,270,000	1,270,000	1,270,000	1,270,000	Banques régionales (plus Bank)	8,378,130	8,366,587	3,012,000	3,819,070	3,817,373	... d'Écosse	8,416,461	8,399,207	3,001,780	3,815,079	3,819,070	... d'Irlande	8,392,240	8,366,617	3,000,240	3,814,000	3,819,070	Totaux	37,077,375	30,186,190	26,361,382	30,147,142	30,360,185
	18 juv. 49	11 nov. 49	4 mai 50	15 juill. 50	22 sept.																																																																																																																																																									
Total	89,371,965	77,899,210	64,946,765	53,218,020	39,299,473																																																																																																																																																									
Billets émis	89,371,965	77,899,210	64,946,765	53,218,020	39,299,473																																																																																																																																																									
Arct. Débits du gouvernement	8,426,893	8,426,893	3,452,800	3,452,800	3,452,800																																																																																																																																																									
Autres débits	14,882,259	13,749,259	13,293,768	13,118,010	13,429,473																																																																																																																																																									
Mouvements d'or et d'argent																																																																																																																																																														
Totaux	89,371,965	77,899,210	64,946,765	53,218,020	39,299,473																																																																																																																																																									
Département des Finances																																																																																																																																																														
Paior. Capital des administrateurs	12,353,000	14,838,000	11,833,000	11,833,000	11,833,000																																																																																																																																																									
Réserves	3,000,000	3,830,000	3,830,000	3,830,000	3,830,000																																																																																																																																																									
Autres... ...de la Banque	14,736,117	13,206,916	13,195,000	13,118,010	13,429,473																																																																																																																																																									
Autres dépôts																																																																																																																																																														
États à payer à 7 jours et moins																																																																																																																																																														
Totaux	16,089,117	17,163,973	17,152,900	17,076,480	17,114,786																																																																																																																																																									
Arct. Généralités... ...de la Banque	19,011,897	20,114,297	17,450,907	19,411,486	19,729,433																																																																																																																																																									
Billets de Banque en poche/bourse	8,426,893	8,426,893	3,452,800	3,452,800	3,452,800																																																																																																																																																									
Mouvements d'or et d'argent																																																																																																																																																														
Totaux	41,006,817	43,681,207	39,439,377	39,446,951	39,299,473																																																																																																																																																									
	18 juv. 49	11 nov. 49	4 mai 50	15 juill. 50	22 sept.																																																																																																																																																									
Total	37,077,375	30,187,458	25,769,510	21,418,452	19,956,740																																																																																																																																																									
Banque d'Angleterre	37,077,375	30,187,458	25,769,510	21,418,452	19,956,740																																																																																																																																																									
Banques provinciales	1,270,000	1,270,000	1,270,000	1,270,000	1,270,000																																																																																																																																																									
Banques régionales (plus Bank)	8,378,130	8,366,587	3,012,000	3,819,070	3,817,373																																																																																																																																																									
... d'Écosse	8,416,461	8,399,207	3,001,780	3,815,079	3,819,070																																																																																																																																																									
... d'Irlande	8,392,240	8,366,617	3,000,240	3,814,000	3,819,070																																																																																																																																																									
Totaux	37,077,375	30,186,190	26,361,382	30,147,142	30,360,185																																																																																																																																																									
Mathematical formulas	<pre><LayoutTag TYPE="Formula" LABEL="MathFormula" DESCRIPTION="4 formulas"/></pre>	<p>On <TextLine> elements</p> <p>On <String> elements</p> <p> $bO = AG - Am ; FO = FG - bm ;$ $bF = \sqrt{bO^2 + FO^2} ; \sin. bFO = \frac{bO}{bF}$ </p> <p>Cela posé, le tétraèdre devant être en équilibre, sous l'action des forces élastiques qui s'exercent sur ses quatre faces, et des forces qui sollicitent sa masse, les sommes des composantes de ces forces, estimées suivant chaque axe, devront être nulles. A la somme des composantes suivant l'axe des x, la face inclinée fournira le terme Xw, la face a, le terme $-N_1 \cdot mw$; la face b, $-T_3 \cdot nw$; la face c, $-T_2 \cdot pw$; les forces qui agissent sur la masse donneront un terme égal à $\rho \cdot X_0$ multiplié par le volume du tétraèdre, qui est un infinité petit du troisième ordre; ce cinquième terme disparaîtra donc à la suite des quatre autres, qui sont des infinités petits du second ordre. Égalant la somme trouvée à zéro, et divisant par w, on obtient la première des équations :</p>																																																																																																																																																												

Chemical formulas (text)	<pre><LayoutTag TYPE="Formula" LABEL="ChemFormula" DESCRIPTION="Ka formula"/></pre>	$K_A = \frac{[A^-]_{eq} [H_3O^+]_{eq}}{[AH]_{eq}}$	
Physics formulas	<pre><LayoutTag TYPE="Formula" LABEL="PhyFormula" DESCRIPTION="e and v formulas"/></pre>	<p>18. Exercices: Les formules qui donnent l'espace et la vitesse dans le mouvement de chute libre, le corps tombant sans vitesse initiale, sont</p> <p>(9) $e = \frac{1}{2} gt^2$, (10) $v = gt$.</p> <p>En éliminant t entre ces deux équations, on a</p> <p>(11) $e = \frac{v^2}{2g}$ où $v = \sqrt{2ge}$,</p> <p>expression utile de la vitesse après une chute d'une hauteur e. Si l'y a une vitesse initiale v_0, les formules deviennent</p>	
Text content under a stamp	<pre><LayoutTag LABEL="TextStamped" DESCRIPTION="PAR"/></pre>		See also "Stamps" in the GraphicalElement section
Handwriting contents, annotations, initials, etc.	<pre><LayoutTag TYPE="Typesetting" LABEL="Handwriting" DESCRIPTION="43g8"/></pre>		Handwriting could also be described with a <GraphicalElement>

	<pre><LayoutTag TYPE="Typesetting" LABEL="Handwriting" DESCRIPTION="N Roret"/></pre>	 A photograph of a historical manuscript page. At the top, the word "AVIS." is printed in a small, bold font. Below it is a block of French text. Underneath the text is a rectangular box containing a handwritten signature, which appears to be "N. Roret".	
Manuscripts	<pre><LayoutTag TYPE="Typesetting" LABEL="Manuscript" DESCRIPTION="Old French manuscript"/></pre>	 A photograph of an Old French manuscript page. The text is written in a dense, dark script. A prominent heading at the top reads "Cy commence la table de ce present livre." Below the heading, there is a large, solid black rectangular redaction box covering several lines of text.	

Script fonts	<pre><LayoutTag TYPE="Typesetting" LABEL="ScriptFonts" DESCRIPTION="Script content"/></pre> <p>font</p>	<p><i>Platinum Gifts for Her Precious Gems Legacy Collection Evile Diamond&Silver</i></p>																																																	
Non-latin fonts	<pre><LayoutTag TYPE="Typesetting" LABEL="NonLatinFont" DESCRIPTION="Chinese ideograms"/>"</pre>	<table border="1"> <thead> <tr> <th colspan="2">INDEX</th> <th>Pages</th> </tr> </thead> <tbody> <tr> <td>Mr. 長</td> <td>Chân 雖 (fils de Gia-Long)</td> <td>43</td> </tr> <tr> <td>An 父 (fils de Nguyễn Phúc-</td> <td>Chêng 父 (appelé aussi</td> <td></td> </tr> <tr> <td>Nguyễn).....</td> <td>Trà 茶.....</td> <td>32</td> </tr> <tr> <td>An 子 (fils de Minh-Mang).</td> <td>Chêng 父.....</td> <td>10</td> </tr> <tr> <td>An-Lang 女郎.....</td> <td>Chêng-Thuong-Vuong.....</td> <td>34</td> </tr> <tr> <td>Ash 桑.....</td> <td>Cô-théak 姑娘.....</td> <td>9, 10</td> </tr> <tr> <td>Bing 霜.....</td> <td>Cv. 婆.....</td> <td>42</td> </tr> <tr> <td>Bíu 婦.....</td> <td>Diên 順 (fils de Nguyễn</td> <td></td> </tr> <tr> <td>Bíu</td> <td>Houng).....</td> <td>17</td> </tr> <tr> <td>Bíu-Crông 寶 銅.....</td> <td>Dia 珍.....</td> <td></td> </tr> <tr> <td>Bíu-Dão 寶 銀.....</td> <td>Dia (fils de Nguyễn Phúc-</td> <td></td> </tr> <tr> <td>Bíu-Lan 寶 銀.....</td> <td>Chá).....</td> <td>31</td> </tr> <tr> <td>Bíu-Liêm 寶 銀.....</td> <td>Duân 雜.....</td> <td>42</td> </tr> <tr> <td>Bíu-Lòi 寶 銀.....</td> <td>Dục 显 (appelé aussi Bíu</td> <td></td> </tr> <tr> <td>etc.</td> <td>寶.....</td> <td>33</td> </tr> </tbody> </table>	INDEX		Pages	Mr. 長	Chân 雖 (fils de Gia-Long)	43	An 父 (fils de Nguyễn Phúc-	Chêng 父 (appelé aussi		Nguyễn).....	Trà 茶.....	32	An 子 (fils de Minh-Mang).	Chêng 父.....	10	An-Lang 女郎.....	Chêng-Thuong-Vuong.....	34	Ash 桑.....	Cô-théak 姑娘.....	9, 10	Bing 霜.....	Cv. 婆.....	42	Bíu 婦.....	Diên 順 (fils de Nguyễn		Bíu	Houng).....	17	Bíu-Crông 寶 銅.....	Dia 珍.....		Bíu-Dão 寶 銀.....	Dia (fils de Nguyễn Phúc-		Bíu-Lan 寶 銀.....	Chá).....	31	Bíu-Liêm 寶 銀.....	Duân 雜.....	42	Bíu-Lòi 寶 銀.....	Dục 显 (appelé aussi Bíu		etc.	寶.....	33	
INDEX		Pages																																																	
Mr. 長	Chân 雖 (fils de Gia-Long)	43																																																	
An 父 (fils de Nguyễn Phúc-	Chêng 父 (appelé aussi																																																		
Nguyễn).....	Trà 茶.....	32																																																	
An 子 (fils de Minh-Mang).	Chêng 父.....	10																																																	
An-Lang 女郎.....	Chêng-Thuong-Vuong.....	34																																																	
Ash 桑.....	Cô-théak 姑娘.....	9, 10																																																	
Bing 霜.....	Cv. 婆.....	42																																																	
Bíu 婦.....	Diên 順 (fils de Nguyễn																																																		
Bíu	Houng).....	17																																																	
Bíu-Crông 寶 銅.....	Dia 珍.....																																																		
Bíu-Dão 寶 銀.....	Dia (fils de Nguyễn Phúc-																																																		
Bíu-Lan 寶 銀.....	Chá).....	31																																																	
Bíu-Liêm 寶 銀.....	Duân 雜.....	42																																																	
Bíu-Lòi 寶 銀.....	Dục 显 (appelé aussi Bíu																																																		
etc.	寶.....	33																																																	

Press content

Masterhead	<p><LayoutTag LABEL="Masterhead" DESCRIPTION="L' AURORE - avril 1942"/></p>		<p>Masterhead could also be described with a structural tag (see section 3.).</p>
------------	---	---	---

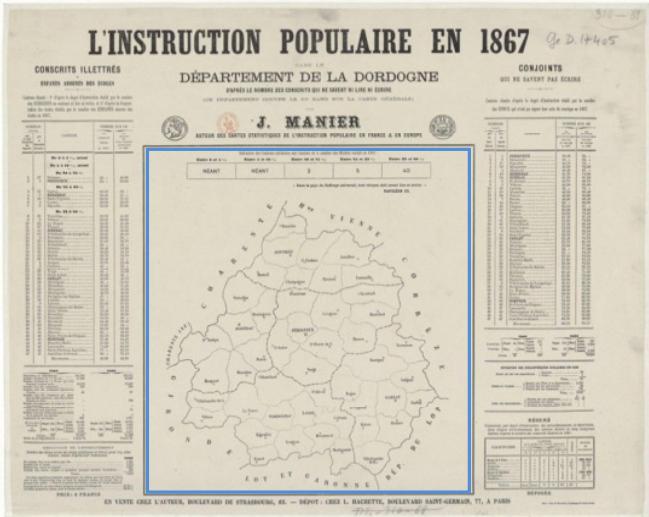
Advertisements,
announcements,
public relation
pages

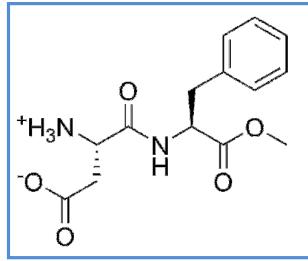
```
<LayoutTag  
LABEL="Advertisements"  
DESCRIPTION="AVIS"/>
```

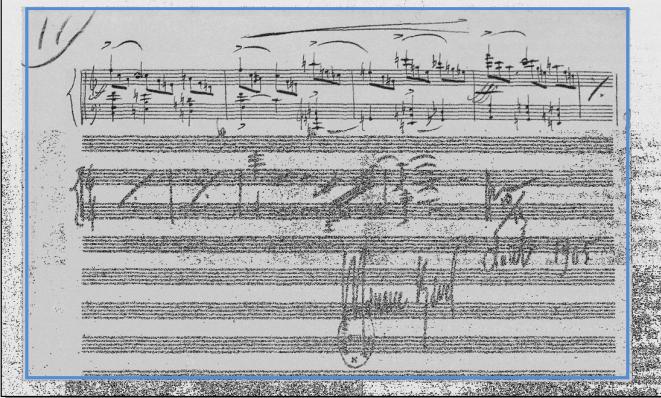
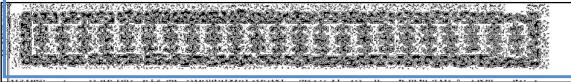
Small ads,
obituaries,
weather forecast,
train tables,
etc.

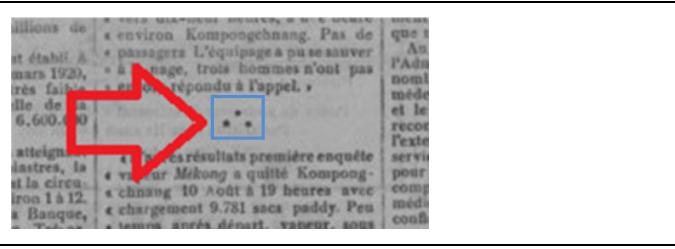
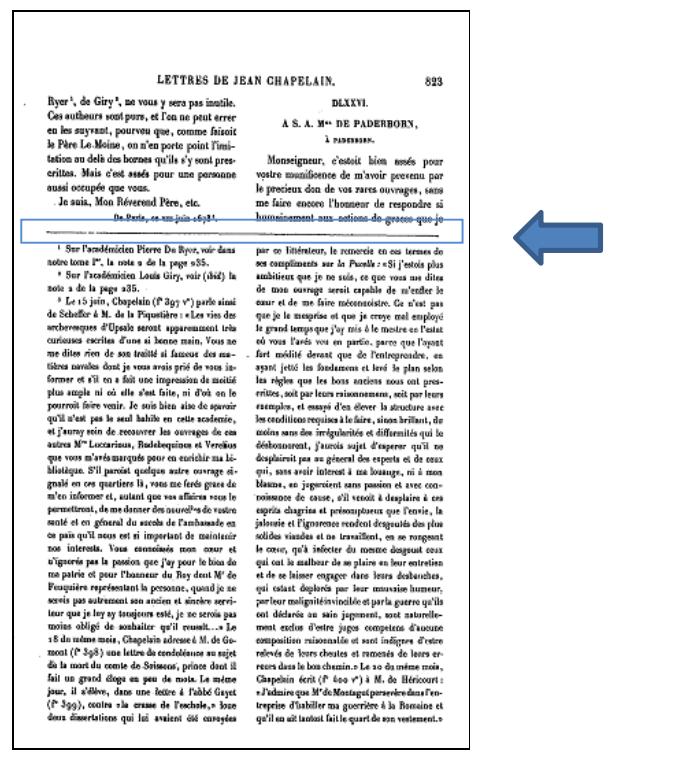
```
<LayoutTag LABEL="SmallAds"  
DESCRIPTION="Petites  
annonces"/>
```

Graphical tags, to be applied on <ComposedBlock> or <Illustration> elements

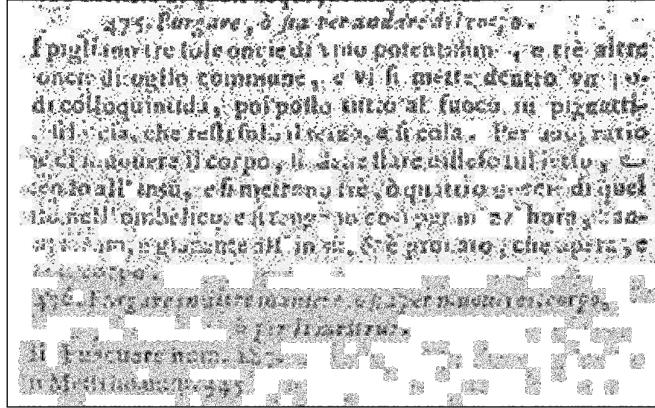
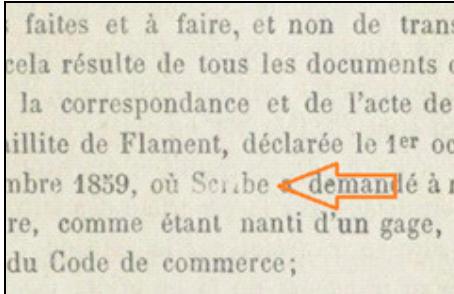
Use cases	ALTO tag	Examples	Remarks
Graphics, charts or graphs, linedrawings, maps, photographs, engravings, etc.	<pre><LayoutTag LABEL="Map"/></pre> <pre><LayoutTag LABEL="Engraving"/></pre>		

	<pre>... <LayoutTag LABEL="Graphic"/> <LayoutTag LABEL="Chart"/> <LayoutTag LABEL="Linedrawing"/> <LayoutTag LABEL="Photo"/> ...</pre>		
Chemical formulas (graphic)	<pre><LayoutTag TYPE="Formula" LABEL="ChemFormula" DESCRIPTION="Aspartame formula"/></pre>		

Musical notations	<pre><LayoutTag LABEL="MusicalScore" DESCRIPTION="Maurice Ravel"/></pre>		
Other graphical tags to be applied on <ComposedBlock>, <GraphicalElement> elements			
Use cases	ALTO tag	Examples	Remarks
Ornaments, tail pieces	no need	 	

Horizontal separator lines between paragraphs	<pre><LayoutTag LABEL="TransitionSep"/></pre>		
Horizontal separator lines between text block and footnotes	<pre><LayoutTag LABEL="FootnoteSep"/></pre>		

Stamps	<LayoutTag LABEL="Stamp" DESCRIPTION="Dépôt légal Vosges 1891"/>	An illustration featuring a large initial 'A' at the top. Below it, a person wearing a hat and holding a umbrella sits on a donkey. The donkey is walking towards the right. In the background, there is a small building or structure. At the bottom, the letters 'A-NE' are visible. A blue rectangular box highlights a circular stamp in the upper left corner of the illustration. The stamp contains text that is partially legible, including 'PELLERIN & C°, imp.-édit.' and 'Dépôt légal Vosges 1891'.	
Dropped initials	<LayoutTag LABEL="DropCap" DESCRIPTION="A"/>	A decorative initial 'S' with intricate scrollwork and floral patterns. To the right of the initial, there is a block of French text: <p>JAN MARTIN SECRETAIRE DE MONSEIGNEUR le Reverendissime Cardinal de Lenoncourt. AUX LECTEURS. I vous desirez, Messieurs, entendre à peu de parolles ce qui est contenu en cest œuvre, sachez que Poliphile dict avoir veu en songe des choses admirables, entre lesquelles il en descrit plusieurs antiques dignes de memoire, comme Pyramides, Obelisques, grandes ruines d'edifices, la difference des colonnes, leurs mesures, piedestalz, bazes, et chapiteaux dont elles sont ornées. Puis les architraves, frizes, cornices, et frontispices avec leurs ouvrages. Un grand cheval, un Elephant de merveilleuse grandeur, un Colosse, et une porte magnifique, avec son plant, ordonnance, moulures, et besongne de taille. Apres comment cinq belles Nymphes le menerent</p>	

Misc, to be applied on any elements			
Use cases	ALTO tag	Examples	Remarks
Illegible contents	<pre><LayoutTag LABEL="Illegible" DESCRIPTION="scan quality problem"/></pre> <pre><LayoutTag LABEL="Illegible" DESCRIPTION="illegible word, may be 'Scribe' "/></pre>	 	On <TextBlock> or <TextLine> elements. On <String> element.

Noise	<LayoutTag LABEL="Noise" DESCRIPTION="scan noise"/>	<p>itable nuée de reporters, non mais de tous les pays du</p> <p>II</p> <p>UN ARTICLE DE PIERRE HUMBOURG, DANS LE « MATIN », — LA SIGNATURE DU VAMPIRE. — LE SOUVENIR DE JACK « L'ÉVENTEUR. — ENCORE UNE FAUSSE PISTE. — LA PIOCHE ET LA PELLE. — UN TYPE ÉTRANGE. — L'HOMME-FEMME. — ATMOSPHÈRE DE LÉGENDE. — LE CROQUIS RÉVÉLATEUR. — LES QUARANTE COMMISSAIRES. — LA FEMME EN SOIE ROSE. — LA GUINGUETTE DU MOULIN. — « NITCHS », — LA FEMME MYSTÉRIEUSE. — LE VAMPIRE ÉCRIT... ÉCRIT... — UN FOU OU UN FARCEUR ?...</p> <p>ourieux de ses informations, de ses lecteurs en envoyant à meilleurs rédacteurs, Pierre</p> <p>« Voici comment le vampire ses crimes. « Les neuf victimes portent tous à la mort la même sentence.</p>	Noise regions: no real data, only artifacts on the document or scanner noise.
Unknown content	<LayoutTag LABEL="Unknown" DESCRIPTION="text?"/>		To be used if the region type cannot be ascertained.

3. Structural tagging

The ALTO format captures the layout and the full text of a page. One intention of OCRing is full text retrieval. But full text retrieval may benefit from marking additional labelling elements such as page headers (running titles), page numbers, signature marks, etc. These are different type of structural elements than intellectual entities which record the intellectual structure of a document (The intellectual structure is not recorded in ALTO, but in a container format such as METS.)

If these elements are labelled, a full text search could either be restricted only to these elements, or the elements could be excluded from a search. A running title that appears on every page could e.g. manipulate the ranking of a resource or deliver invalid hits. Other labelled information such as the page number could be used for quality assurance purpose, document navigation, etc.

This structural tagging can be implemented in a mass digitisation workflow to extract structural or functional features (TOC entries, headings, headers, footnotes, article headings, etc.) or as an editing and correction facility for the improvement of already digitised books. Apart from METS/ALTO, enhanced PDFs and ebooks can also be generated if some structural information is available.

XML schema

ALTO tag element: <StructureTag>

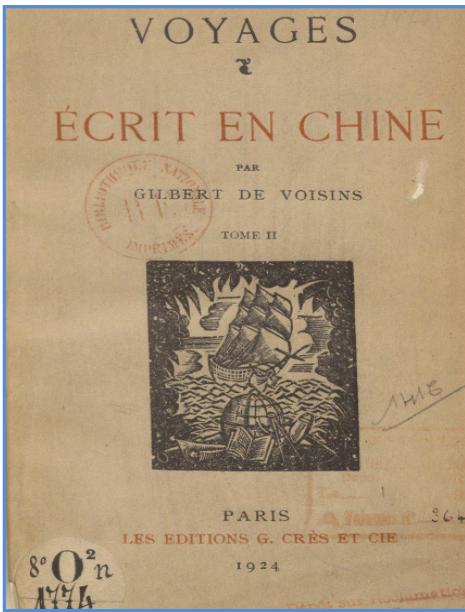
Tag attributes:

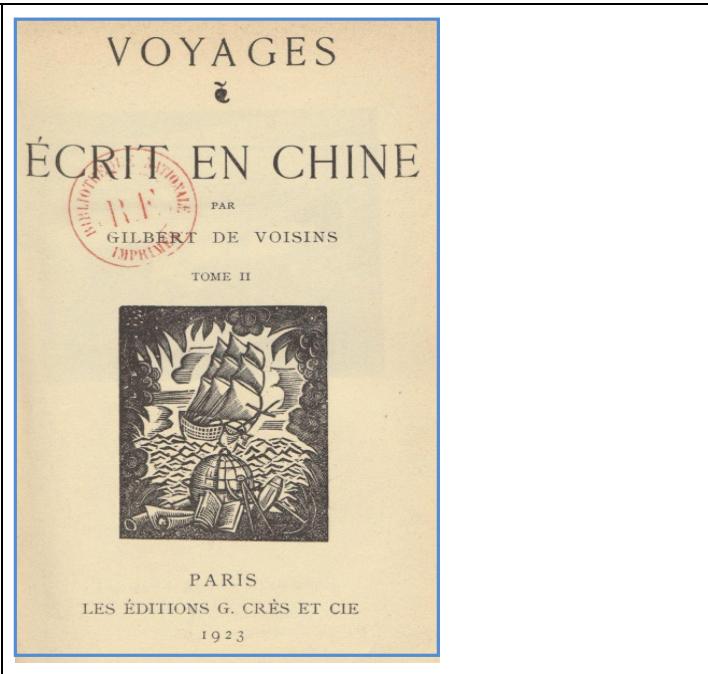
- **ID**: Tag id
- **LABEL**: Structural tag
- **TYPE** (optional): Supertype or category of the tag
- **DESCRIPTION** (optional): Description of the tag, description of the content, or the content itself
- **XmlData** (optional): Any XML encoded metadata helping to describe the tag or the vocabulary used to describe the tag.
- **URI** (optional): Any authority URI relevant for the tag element

Good practices, limitations

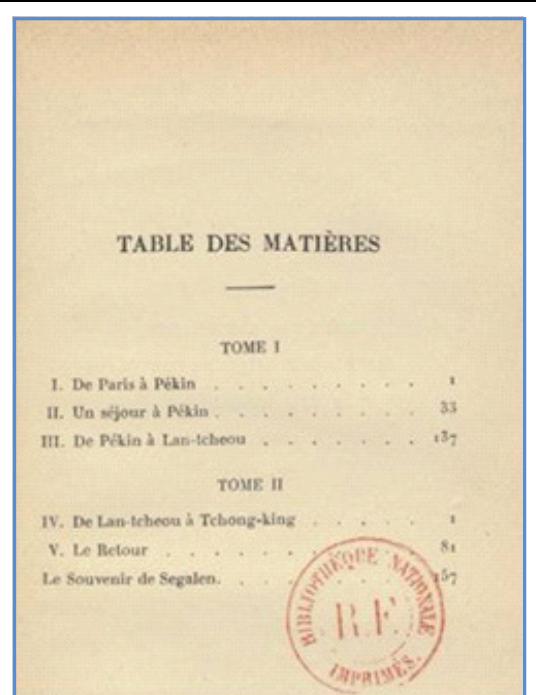
- Structural tagging in ALTO can be considered as awkward (compared to the classic and recommended METS-ALTO solution), as ALTO does not foresee the storing of any semantic information. But it could be useful:
- If a METS container is not used in the digitization project.
- If a simple and limited capturing of structural information is needed.
- Besides, tagging in ALTO has some benefits: engineering simplicity (compared to a METS-ALTO implementation), autonomy (content and structure are stored in a unique file).

Text tags, to be applied on <ComposedBlock>, <TextBlock>, <TextLine> or <String> elements

Use Cases	ALTO tag	Examples	Remarks
Covers	<pre><StructureTag TYPE="Functional" LABEL="Cover"/> ... <StructureTag TYPE="Functional" LABEL="cover"> <XmlData> <Scheme>epub</Scheme> </XmlData> </RoleTag></pre>		<p>http://gallica.bnf.fr/ark:/12148/bpt6k6207567r/f1.planchecontact</p> <p>Description of the vocabulary used.</p>

Title pages	<StructureTag TYPE="Functional" LABEL="TitlePage"/>	 The title page of a book titled "VOYAGES ÉCRIT EN CHINE". The author is listed as "GILBERT DE VOISINS". The publisher is "LES ÉDITIONS G. CRÈS ET CIE". The year of publication is "1923". A red circular stamp is visible on the left side of the page. The page is framed by a blue border.	
-------------	---	--	--

Frontmatter (foreword, prologue, notice to the reader, warning, advertisements, copyright page, publisher catalogs, etc.)	<pre><StructureTag TYPE="Functional" LABEL="Catalog"/> ... <StructureTag TYPE="Functional" LABEL="CopyrightPage"/> <StructureTag TYPE="Functional" LABEL="Foreword"/> <StructureTag TYPE="Functional" LABEL="Notice"/> ...</pre>	<p>DU MÊME AUTEUR</p> <p><i>La petite Angoisse</i>, roman. <i>Pour l'Amour du Laurier</i>, roman. <i>Le Démon secret</i>, roman. <i>Sentiments</i>, critique. <i>Les Moments perdus de John Shag</i>. <i>Le Bar de la Fourche</i>, roman. <i>L'Enfant qui prit peur</i>, roman. <i>Le Mirage</i>, roman. <i>L'Esprit impur</i>, roman. <i>Fantastiques</i>, petits poèmes. <i>La Conscience dans le Mal</i>, roman.</p> <p>Prochainement.</p> <p><i>Le Jour naissant</i>.</p>	
---	---	---	--

Tables of content	<StructureTag TYPE="Functional" LABEL="TOC"/>	 <p>TABLE DES MATIÈRES</p> <p>—</p> <p>TOME I</p> <p>I. De Paris à Pékin 1 II. Un séjour à Pékin 33 III. De Pékin à Lan-icheou 137</p> <p>TOME II</p> <p>IV. De Lan-icheou à Tchong-king 1 V. Le Retour 81 Le Souvenir de Segalen. 57</p> <p>BIBLIOTHÈQUE NATIONALE REF. IMPRIMERIE</p>	Global tables of content Internal tables of content (chapter beginnings, etc.)
-------------------	---	--	---



Body matter	<pre><StructureTag TYPE="Structural" LABEL="BodyMatter"/></pre>	<p>Lan-tcheou. 23 octobre.</p> <p>M'y voilà donc, dans ce canton lointain que je m'étais fixé comme but! Voulant rêver d'aventures, j'avais ouvert mon atlas sur le plus grand des continents et, par jeu, tout au hasard, j'avais piqué la page, parmi les fleuves, les villes, les montagnes...</p> <p>Je regardai le petit point noir.</p> <p>« Heureux, pensai-je, qui peut se trouver là, avant vécu dans le bruit que font les hommes de son pays! ».</p> <p>Sans cesse, j'y songeais, ne quittant plus la carte des yeux et revenant toujours à ce petit</p>	Beginning of the main content
-------------	---	---	-------------------------------

<p>Backmatter (afterword, appendix, illustrations list, tables list, conclusion, glossary, bibliography, colophon, etc.)</p>	<pre><StructureTag TYPE="Functional" LABEL="LOI"/> <StructureTag TYPE="Functional" LABEL="Appendix"/> <StructureTag TYPE="Functional" LABEL="LOT"/> <StructureTag TYPE="Functional" LABEL="Conclusion"/> <StructureTag TYPE="Functional" LABEL="Glossary"/> <StructureTag TYPE="Functional" LABEL="Bibliography"/> ...</pre>	<p style="text-align: center;">TABLE DES ILLUSTRATIONS</p> <table border="0" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; width: 80%;"></th><th style="text-align: right; font-weight: normal;">pages</th></tr> </thead> <tbody> <tr> <td style="vertical-align: top;">TOME I</td><td></td></tr> <tr> <td>Esplanade du Temple du Ciel</td><td style="text-align: right;">32</td></tr> <tr> <td>Temple de Pi-jun-sseu'</td><td style="text-align: right;">48</td></tr> <tr> <td>Muraille nord de la ville tartare</td><td style="text-align: right;">80</td></tr> <tr> <td>Stèle au Si-ing</td><td style="text-align: right;">96</td></tr> <tr> <td>Nos Gens</td><td style="text-align: right;">144</td></tr> <tr> <td>Office de Ou-tai-chan</td><td style="text-align: right;">160</td></tr> <tr> <td>L'arrivée à Ho-tcheou</td><td style="text-align: right;">176</td></tr> <tr> <td>Un pont sur le Fen-ho</td><td style="text-align: right;">192</td></tr> <tr> <td>Dans le loess du Chenoi</td><td style="text-align: right;">208</td></tr> <tr> <td>L'arrivée à Ping-leang</td><td style="text-align: right;">224</td></tr> <tr> <td>Dans le loess du Kansou.</td><td style="text-align: right;">228</td></tr> <tr> <td colspan="2" style="height: 20px;"></td></tr> <tr> <td style="vertical-align: top;">TOME II</td><td></td></tr> <tr> <td>Arrivée à Lan-tchéou</td><td style="text-align: right;">1</td></tr> <tr> <td>Une caravane dans le Kansou.</td><td style="text-align: right;">32</td></tr> <tr> <td>Un pont sur le Kei-chouei</td><td style="text-align: right;">48</td></tr> <tr> <td>Dans les rapides du Ta-ho</td><td style="text-align: right;">48</td></tr> <tr> <td>L'entrée souterraine d'Omei-chan.</td><td style="text-align: right;">64</td></tr> <tr> <td>La jonque de Lohéac</td><td style="text-align: right;">96</td></tr> <tr> <td>Le temple de Che-pao-tchâï</td><td style="text-align: right;">112</td></tr> <tr> <td>Une jonque à double sao.</td><td style="text-align: right;">128</td></tr> <tr> <td>Jonques à Canton</td><td style="text-align: right;">144</td></tr> <tr> <td>Victor Segalen en Chine</td><td style="text-align: right;">156</td></tr> </tbody> </table>		pages	TOME I		Esplanade du Temple du Ciel	32	Temple de Pi-jun-sseu'	48	Muraille nord de la ville tartare	80	Stèle au Si-ing	96	Nos Gens	144	Office de Ou-tai-chan	160	L'arrivée à Ho-tcheou	176	Un pont sur le Fen-ho	192	Dans le loess du Chenoi	208	L'arrivée à Ping-leang	224	Dans le loess du Kansou.	228			TOME II		Arrivée à Lan-tchéou	1	Une caravane dans le Kansou.	32	Un pont sur le Kei-chouei	48	Dans les rapides du Ta-ho	48	L'entrée souterraine d'Omei-chan.	64	La jonque de Lohéac	96	Le temple de Che-pao-tchâï	112	Une jonque à double sao.	128	Jonques à Canton	144	Victor Segalen en Chine	156	
	pages																																																				
TOME I																																																					
Esplanade du Temple du Ciel	32																																																				
Temple de Pi-jun-sseu'	48																																																				
Muraille nord de la ville tartare	80																																																				
Stèle au Si-ing	96																																																				
Nos Gens	144																																																				
Office de Ou-tai-chan	160																																																				
L'arrivée à Ho-tcheou	176																																																				
Un pont sur le Fen-ho	192																																																				
Dans le loess du Chenoi	208																																																				
L'arrivée à Ping-leang	224																																																				
Dans le loess du Kansou.	228																																																				
TOME II																																																					
Arrivée à Lan-tchéou	1																																																				
Une caravane dans le Kansou.	32																																																				
Un pont sur le Kei-chouei	48																																																				
Dans les rapides du Ta-ho	48																																																				
L'entrée souterraine d'Omei-chan.	64																																																				
La jonque de Lohéac	96																																																				
Le temple de Che-pao-tchâï	112																																																				
Une jonque à double sao.	128																																																				
Jonques à Canton	144																																																				
Victor Segalen en Chine	156																																																				

Index	<StructureTag TYPE="Functional" LABEL="Index"/>	<p style="text-align: center;">P</p> <p style="text-align: center;">(suite)</p> <p>PIA DE TOLOMEI (la) : <i>Amour</i>, I, 132, 134 ; <i>Promenades</i>, II, 41. PIA (porta) : <i>Promenades</i>, I, 24, 127, 256 ; II, 54, 66 ; III, 260, 324, 329 ; <i>Idées</i>, 295 ; <i>Correspondance</i>, IX, 329 ; X, 149. PIACENZA : <i>Journal</i>, V, 293. PIANONI : <i>Pages</i>, 56. PIANCANI (Vincent) : <i>Politique</i>, II, 316 ; <i>Table</i>, I, 38, 169. PIAN D'ERBA : <i>Rome</i>, I, 114. PIAN DI SPILLE (Tour de) : <i>Correspondance</i>, X, 262. PIANORO (peintre) : <i>Ecoles</i>, III, 363, 364. PIANORO (géogr.) : <i>Journal</i>, IV, 326, 327. PIANSANO : <i>Correspondance</i>, X, 206. Pianto d'Armonia (II) : Rossini, I, 64 ; II, 237, 245, 329. PIAT-DESVIALS (Mme) : <i>Brulard</i>, I, 127. PIATTI : <i>Correspondance</i>, X, 206. PIAVE (la) : <i>Napoléon</i>, II, 313 ; <i>Ecoles</i>, II, 125. PIAZZA (peintre) : <i>Ecoles</i>, II, 170, 171.</p>	
Running titles	<StructureTag TYPE="Functional" LABEL="RunningTitle"/>	 <p>LE VAMPIRE DE BURGUNDO</p> <p>Illustration of a scene from Burgundy, showing two figures in a dark, atmospheric setting.</p>	

Volumes, parts, chapters	<pre><StructureTag TYPE="Structural" LABEL="Part" DESCRIPTION="I"/></pre> <pre><StructureTag TYPE="Structural" LABEL="Chapter" DESCRIPTION="I"/></pre>	<p>The image shows the title page of a book. The title is "PHYSIOLOGIE CATHOLIQUE ET PHILOSOPHIQUE". Below it is a decorative horizontal line. A blue rectangular box contains the text "PREMIÈRE PARTIE" and "DE LA VIE DE RELATION". Another blue rectangular box below it contains "CHAPITRE I" and "DES FONCTIONS SENSORIALES OU DES SENS, C'EST-A-DIRE DES SENSATIONS.". The text "Réflexions préliminaires." is written in a smaller font. The main text begins with: "Dominateur du globe et des éléments, l'homme doit entretenir avec toute la nature les relations nécessaires à son existence et à son bien-être; il est donc important qu'il apprécie ces rapports, afin qu'il recherche et attire les objets qu'il a intérêt de connaître, et qu'il fui et repousse ceux qui lui nuisent. Les sens sont des sentinelles avancées placées aux limites de l'existence, lesquelles, à l'aide de la correspondance active et presque incessante des nerfs, transmettent à la souveraine assise sur son trône tout ce qui se passe à l'étranger, c'est-à-dire dans le monde extérieur. Or cette souveraine, c'est l'âme; son trône ou son palais, c'est le cerveau, qui est en même temps le centre de son gouvernement (1). (1) Que l'on ne pense pas que nous prétendions assigner à l'âme un siège matériel et anatomique. Voici à ce sujet ce que nous avons dit dans un autre ouvrage.</p>	
-----------------------------	---	---	--

Titles, subtitles	<pre><StructureTag TYPE="Structural" LABEL="FullTitle"/></pre> <pre><StructureTag TYPE="Structural" LABEL="Title1"/></pre>	<p style="text-align: center;"> Les XXII Lames Hermétiques DU TAROT DIVINATOIRE </p> <hr/> <p style="text-align: center;">INTRODUCTION</p> <p style="text-align: center;">L'Astrologie est un traité de la Providence. ST JÉRÔME.</p> <p style="text-align: center;">Le Tarot, du sanscrit TAR-O, <i>étoile fixe</i> (probablement la polaire qui dans l'astrologie antique symbolisait la tradition immuable) n'est autre chose que la synthèse théosophique et symbolique du dogme pris au milieu des Religions en même temps qu'une méthode simplifiée d'Astrologie, retrouvées par le Mage Hermès (1) surnommé trismégiste qui était hiérophante des Temples de Thèbes, 2000 ans avant J.-C., où il se servait du Tarot comme instrument sacré ; il était alors gravé sur 22 lames d'or, qui portaient en plus des inscriptions en écriture hiéroglyphique et en lettres égyptiennes. L'auteur de la <i>Table d'Emerude</i> et du <i>Pymander</i>,</p>	
-------------------	---	--	--

```
<StructureTag
  TYPE="Structural"
  LABEL="FullTitle"/>
```

```
<StructureTag
  TYPE="Structural"
  LABEL="Title1"
  DESCRIPTION="I"/>

<StructureTag
  TYPE="Structural"
  LABEL="Title2"
  DESCRIPTION="I"/>
```

```
<StructureTag
  TYPE="Structural"
  LABEL="Title2"
  DESCRIPTION="II"/>
```



Numbered titles

Footnote references	<pre><StructureTag TYPE="Functional" TYPE="FootnoteReference" DESCRIPTION="1"/></pre>	<p>Le Tarot, du sanscrit TAR-O, <i>étoile fixe</i> (probablement la polaire qui dans l'astrologie antique symbolisait la tradition immuable) n'est autre chose que la synthèse théosophique et symbolique du dogme primal des Religions en même temps qu'une méthode simplifiée d'Astrologie, retrouvées par le Mage Hermès (1) surnommé trismégiste qui était hiérophante des Temples de Thèbes, 2000 ans avant J.-C, où il se servait du Tarot comme instrument augural ; il était alors gravé sur 22 lames d'or, qui portaient en plus</p> <p>I. L'auteur de la Table d'Emeraude et du Pymandre.</p>	
Footnotes	<pre><StructureTag TYPE="Functional" LABEL="Footnote" DESCRIPTION="1"/></pre>	<p>Le Tarot, du sanscrit TAR-O, <i>étoile fixe</i> (probablement la polaire qui dans l'astrologie antique symbolisait la tradition immuable) n'est autre chose que la synthèse théosophique et symbolique du dogme primal des Religions en même temps qu'une méthode simplifiée d'Astrologie, retrouvées par le Mage Hermès (1) surnommé trismégiste qui était hiérophante des Temples de Thèbes, 2000 ans avant J.-C, où il se servait du Tarot comme instrument augural ; il était alors gravé sur 22 lames d'or, qui portaient en plus</p> <p>I. L'auteur de la Table d'Emeraude et du Pymandre.</p>	
References footnote	to	<pre><StructureTag TYPE="Reference" TYPE="ReferenceToFootnote" DESCRIPTION="1"/></pre>	

Marginalias	<pre><StructureTag TYPE="Functional" LABEL="Marginalia"/></pre>	<p style="text-align: center;">D E L U M I N E. 69</p> <p>riores Planetæ pro cādem ratione non magis à Sole illuminati sint, quām superiores, neque etiam quām Cometæ, qui citra omnem controversiam ab illo remotiores sūnt.</p> <p>Simile quid in vero etiam Mundo obtainere experientia evincit, & tamen non credo, rationem ejus reddi posse, supposito lumen aliud quid esse in objectis, quām ejusmodi actionem aut dispositiōnem, qualem expolui. Actionem dico, aut dispositionem; observato enim illo quod jam dedi demonstratum, quod * vacuo existente spatio, in quo Sol est, coeli ejusdem partes versus oculos spe- ctantium tenderent cādem ratione, quām cū ab illius materiā im- pelluntur, & æquali ferme cum impetu; percipi quidem poterit, haud opus esse, ut aliquam in se actionem habeat, neque etiam ut quid aliud sit, quām merum spatiū, ut qualem ipsum videmus ap-</p> <div style="text-align: right; margin-top: -20px;"> Part. 3. Princ. §. 64 </div>	
Figure captions	<pre><StructureTag TYPE="Functional" LABEL="FigureCaption"/></pre>	 <p>Above the illustration, the title 'LE VAMPIRE DU DUNSETHORPE' is printed in a decorative font. Below the illustration, a caption reads: 'Illustration de deux de nos lecteurs, ayant vu dans une grotte des ours sauvages dormir. (page 15.)'</p>	

Figure references	<pre><StructureTag TYPE="Functional" LABEL="FigureReference" DESCRIPTION="9"/></pre>		
References to figure	<pre><StructureTag TYPE="Reference" LABEL="ReferenceToFigure" DESCRIPTION="9"/></pre>		

Table captions	<pre><StructureTag TYPE="Functional" LABEL="TableCaption"/></pre>	<p>2. — Régime fiscal applicable aux fonds d'États étrangers et aux sociétés étrangères non abonnées.</p> <p>Nous avons indiqué les raisons pour lesquelles le régime de l'abonnement n'a pas paru applicable aux titres d'emprunts émis par les gouvernements étrangers. Par ailleurs, l'abonnement laisse en dehors de son domaine d'application les titres de sociétés et collectivités étrangères dont l'introduction en France n'est pas le fait personnel desdites sociétés et collectivités. Il convient donc d'étudier quel est le régime fiscal auquel sont assujettis les titres de ces deux catégories circulant en France.</p> <p><i>a. — FONDS D'ÉTATS ÉTRANGERS.</i></p> <p>Les fonds d'États étrangers ne sont pas assujettis à la formalité préalable de l'abonnement pour pouvoir être négociés en France. Mais ils sont assujettis à l'impôt du timbre et à l'impôt sur le revenu.</p> <p><i>e. — Impôt du timbre.</i></p> <p>En vertu des dispositions législatives rappelées par l'article 19 du décret de codification du 29 décembre 1926, les titres de rentes, emprunts et autres effets publics des gouvernements étrangers sont soumis à un droit de timbre au comptant, calculé en principe sur leur valeur nominale. Néanmoins, pour ceux de ces titres qui sont cotés à la bourse officielle et dont le cours moyen pendant l'année précédente est tombé au-dessous de la moitié du pair, la perception s'effectue sur la valeur négociable déterminée par ce cours.</p> <p>TAUX DU DROIT. — Les taux successifs du droit de timbre sont indiqués par le tableau ci-dessous :</p> <table border="1" data-bbox="882 878 1389 1090"> <tbody> <tr> <td>Loi du 18 mai 1863.....</td> <td>0,50 0/0</td> </tr> <tr> <td>— 8 juin 1864.....</td> <td>1 —</td> </tr> <tr> <td>— 25 mai 1872.....</td> <td>0,75 pour 500 fr. et au-dessous. 1,50 de 500 fr. à 1.000 fr. 3 de 1.000 fr. à 2.000 fr. 4,50 de 2.000 fr. à 3.000 fr.</td> </tr> <tr> <td>Loi du 28 décembre 1895.....</td> <td>0,50 0/0</td> </tr> <tr> <td>— 13 avril 1898.....</td> <td>1 —</td> </tr> <tr> <td>— 30 janvier 1907.....</td> <td>2 —</td> </tr> <tr> <td>— 30 juillet 1913.....</td> <td>3 —</td> </tr> <tr> <td>— 29 mars 1914.....</td> <td>2 —</td> </tr> <tr> <td>— 24 mars 1924.....</td> <td>2,40 —</td> </tr> <tr> <td>— 3 août 1926.....</td> <td>4 —</td> </tr> </tbody> </table>	Loi du 18 mai 1863.....	0,50 0/0	— 8 juin 1864.....	1 —	— 25 mai 1872.....	0,75 pour 500 fr. et au-dessous. 1,50 de 500 fr. à 1.000 fr. 3 de 1.000 fr. à 2.000 fr. 4,50 de 2.000 fr. à 3.000 fr.	Loi du 28 décembre 1895.....	0,50 0/0	— 13 avril 1898.....	1 —	— 30 janvier 1907.....	2 —	— 30 juillet 1913.....	3 —	— 29 mars 1914.....	2 —	— 24 mars 1924.....	2,40 —	— 3 août 1926.....	4 —	
Loi du 18 mai 1863.....	0,50 0/0																						
— 8 juin 1864.....	1 —																						
— 25 mai 1872.....	0,75 pour 500 fr. et au-dessous. 1,50 de 500 fr. à 1.000 fr. 3 de 1.000 fr. à 2.000 fr. 4,50 de 2.000 fr. à 3.000 fr.																						
Loi du 28 décembre 1895.....	0,50 0/0																						
— 13 avril 1898.....	1 —																						
— 30 janvier 1907.....	2 —																						
— 30 juillet 1913.....	3 —																						
— 29 mars 1914.....	2 —																						
— 24 mars 1924.....	2,40 —																						
— 3 août 1926.....	4 —																						
Table references	<pre><StructureTag TYPE="Functional" LABEL="TableReference" DESCRIPTION="1" /></pre>																						
References table to	<pre><StructureTag TYPE="Reference" LABEL="ReferenceToTable" DESCRIPTION="1" /></pre>																						

Page numbers	<pre><StructureTag TYPE="Functional" LABEL="PageNumber" DESCRIPTION="937"/></pre>		
Reference to page	<pre><StructureTag TYPE="Reference" LABEL="ReferenceToPage" DESCRIPTION="8"/></pre>	<p><i>Nous sortions de chez le boulanger, quand nous vîmes devant nous, une femme qui nous regardait fixement. (Page 8).</i></p> <p style="text-align: center;">— 9 —</p>	
Bullets lists, numbered lists	<pre><StructureTag TYPE="Functional" LABEL="UL"/> <StructureTag TYPE="Functional" LABEL="OL"/></pre>		

	Press content	
Masterhead, imprint	<pre><StructureTag TYPE="Functional" LABEL="Masterhead"/></pre>	
Newspaper title	<pre><StructureTag TYPE="Functional" LABEL="FullTitle" DESCRIPTION="THE WINCHESTER NEWS"/></pre>	
Issue date	<pre><StructureTag TYPE="Functional" LABEL="Date" DESCRIPTION="October 31, 1910"/></pre>	

<p>Section headings, rubrics, column titles, etc.</p> <pre> <StructureTag TYPE="Structural" LABEL="SectionHeading" DESCRIPTION="SOCIETY"/> </pre>	<p>http://chroniclingamerica.loc.gov/lccn/sn86069133/1910-10-31/ed-1/seq-3/</p>
---	--

<p>Advertisements, announcements, public relation pages</p>	<pre><StructureTag TYPE="Structural" LABEL="SectionHeading" DESCRIPTION="Advertisements" /></pre>		
<p>Individual content item of advertisements, announcements, public relation pages</p>	<pre><StructureTag TYPE="Structural" LABEL="Heading" DESCRIPTION="Advertisement" /></pre>		

	...		
Location, spatial information, place name	<p><StructureTag TYPE="Functional" LABEL="Place" DESCRIPTION="CLAY CITY, Ky"/></p>	<p>GREAT CROWD HEARS BYRD</p> <p>Democratic Nominee For Congress and R. R. Friend, of Irvine, Speak to Powell County Voters.</p> <p>CLAY CITY, Ky., Oct. 31.—Hon. A. Floyd Byrd, Democratic nominee for Congress from the Tenth Congressional District and Robert R. Friend, of Irvine, Ky., Saturday ad-</p> <p>her injuries was made an eye specialist, at tation, announced the injured and the sig but said he did not th feel the child's appe</p> <p>Mr. Norris, cattle last week and made among them 350 he Bigstaff, average weight at 6 cents; 150 of average weight 1.40 cents, and 60 of Hennage weight 1.450 pon These cattle were bou timore market, as we Morris purchased. M purchased of Ben F.</p>	Coverage note about a content
Date, dateline	<p><StructureTag TYPE="Functional" LABEL="Date" DESCRIPTION="Oct. 31."/></p>	<p>GREAT CROWD HEARS BYRD</p> <p>Democratic Nominee For Congress and R. R. Friend, of Irvine, Speak to Powell County Voters.</p> <p>CLAY CITY, Ky., Oct. 31.—Hon. A. Floyd Byrd, Democratic nominee for Congress from the Tenth Congressional District and Robert R. Friend, of Irvine, Ky., Saturday ad-</p> <p>her injuries was made an eye specialist, at tation, announced the injured and the sig but said he did not th feel the child's appe</p> <p>Mr. Norris, cattle last week and made among them 350 he Bigstaff, average weight at 6 cents; 150 of average weight 1.40 cents, and 60 of Hennage weight 1.450 pon These cattle were bou timore market, as we Morris purchased. M purchased of Ben F.</p>	Coverage note about a content

	<p>Article authors, copyright note, etc.</p> <pre><StructureTag TYPE="Functional" LABEL="ArticleAuthor" DESCRIPTION="Léon Lafage"/></pre>		<p>http://gallica.bnf.fr/ark:/12148/bpt6k297360g/f1.iimage</p> <p>Authors could also be tagged with a <code><RoleTag></code> (see section 5.).</p>
Illustration	<pre><StructureTag TYPE="Functional" LABEL="Illustration" DESCRIPTION="John W. Langley"> <XmlData> <!-- bibliographic metadatas of the photography --> </XmlData> <StructureTag/></pre> <pre><StructureTag TYPE="Functional" LABEL="Caption"/></pre>		<p>http://chroniclingamerica.loc.gov/lccn/sn86069133/1910-10-31/ed-1/seq-6/</p> <p>If the intellectual value of an illustration is high, it can be captured with a structural tag.</p>

4. Named Entities tagging

Named-entity recognition (NER) is a subtask of information extraction that seeks to locate and classify atomic textual elements into predefined categories (names of persons, organizations, locations, expressions of times, quantities, etc).

Representing the result of a NER processing in ALTO has several pros:

- Simplicity: no need to use another format for NE storage, no need to make links between this format and the ALTO block coordinates; no need to use a METS-ALTO approach.
- Autonomy: contents and NE are stored in the same file.

XML schema

ALTO tag element: <NamedEntityTag>

Tag attributes:

- **ID**: Tag id
- **LABEL**: Named entity category (Person, Organization, Location, etc.)
- **TYPE** (optional): Supertype or category of the named entity type
- **DESCRIPTION** (optional): Named entity textual content
- **Xmldata** (optional): Any XML encoded metadata helping to describe the tag or the vocabulary used to describe the tag
- **URI** (optional): Any authority URI relevant for the tag element: NE authority repositories, gazetteers, etc.

Good practices, limitations

- Multiple authorities can be assigned to a named entity (see example bellow).
- Nested tags can be described (see example bellow).
- Occurrences of the same NE can't clearly be described if they are near one of the other. The first two strings reference the "JS Bach" NE and the last three reference the same "JS Bach" NE, but the mechanism can't say where is the frontier.

... JS Bach, (Johann Sebastian Bach) ...

Text tags, to be applied on <ComposedBlock>, <TextBlock>, <TextLine> or <String> elements			
Use Cases	ALTO tag	Examples	Remarks
Locations	<pre> <!--Simple tagging --> <NamedEntityTag LABEL="Location" DESCRIPTION="Lexington"/> <!-- URI Authority --> <NamedEntityTag LABEL ="Location" DESCRIPTION="Lexington" URI="http://www.geonames.org/4941935" /> <!-- Extra attributes --> <NamedEntityTag LABEL ="Location" DESCRIPTION="Kentucky"> <XmlData> <NEC>1.0</NEC> <!-- NE confidence value --> <Variants>Kenekuke Kentaki Kentákii Hahoodzo Kéntòki Kentórkj Kentuki Kentukia Kentukio Kentukis Kentukki Kentukki Kéntukki Shitati Khén-thap-kí Kentaki کنتاکی گیٹنگ کentákí Kentaki Kentakki Kentukki Kentýkki Кінтааки Штат Kentuki Ченсунмұхаммед әбділұзо кеңткі کentکی کentکی کentکی کentکی گentتکی کentکی کentکی </Variants> </XmlData> </NamedEntityTag> <!-- Multiple Authorities --> </pre>		http://chroniclingamerica.loc.gov/lcn/sn86069133/1910-10-31/ed-1/seq-1/ http://chroniclingamerica.loc.gov/lcn/sn86069133/1910-10-31/ed-1/seq-1/ocr.xml

	<pre> <NamedEntityTag ID="15a" LABEL="Location" DESCRIPTION="Louisville" URI="http://www.geonames.org/4299276" /> <NamedEntityTag ID="15b" LABEL="Location" DESCRIPTION="Louisville" URI="mygeonames:louisville" /> ... <String CONTENT="Louisville" WC="1.0" TAGREFS="15a 15b"> </pre>		
Persons	<pre> <!-- Simple tagging --> <NamedEntityTag LABEL="Person" DESCRIPTION="Dr. Reynolds" /> <!-- URI Authority --> <NamedEntityTag LABEL="Person" DESCRIPTION="Mr. Norris" URI="http://catalogue.bnf.fr/servlet/autori te?ID=11916925" /> <!-- embedded authority description --> <NamedEntityTag LABEL="Person" DESCRIPTION="James M Bigstaff " > <XmlData> <MADS xmlns="http://www.loc.gov/mads/"> <authority> <name type="personal" authority="naf"> <namePart>Bigstaff, James M</namePart> <namePart type="date">1835- 1882</namePart> </name> </authority> </pre>	<p>her injuries was made. Dr. Reynolds, an eye specialist, after the examination, announced the eye permanently injured and the sight forever gone, but said he did not think it would affect the child's appearance.</p> <p>Mr. Norris, cattle buyer, was here last week and made many purchases, among them 350 head of James M. Bigstaff, average weight 1,500 pounds at 6 cents; 150 of J. T. Highland, average weight 1,400 pounds, at 6 cents, and 60 of Henry C. Hall, average weight 1,450 pounds at 6 cents. These cattle were bought for the Baltimore market, as were all the cattle Morris purchased. Mr. W. B. Greene purchased of Ben F. Mark 40 cattle, average 1,400 pounds, at the prevailing prices. There are yet about 1,000 cattle for the export market unsold in this county.</p>	

	</MADS> </XmlData></NamedEntityTag>		
Organization s	<pre> <!--Simple tagging --> <NamedEntityType ID="O20" LABEL="Organization" DESCRIPTION="Central Kentucky Tobacco Warehouse Company"/> <!-- Nested tags --> <NamedEntityType ID="L11" LABEL="Location" DESCRIPTION="Kentucky"/> ALTO content: ... <String CONTENT="Central" TAGREFS="O20"> </TextLine> <TextLine> <String CONTENT="Kentucky" TAGREFS="O20 L11"/> <String CONTENT="Tobacco" TAGREFS="O20"> <String CONTENT="Warehouse" TAGREFS="O20"/> <String CONTENT="Company" TAGREFS="O20"/> ... </pre>	<p style="text-align: center;">WAREHOUSE SOLD.</p> <p>LEXINGTON, Ky., Oct. 31.—The biggest business transaction engineered in Lexington in many months will be completed Monday when the final papers in the sale of the Central Kentucky Tobacco Warehouse Company will be signed, according to a well-authenticated story current Saturday night. The deal involves an exchange of from \$95,000 to \$100,000, and has been under consideration for several weeks past.</p> <p>The purchasers of this building which is the largest tobacco warehouse in the world, is the Louisville Tobacco Warehouse Company of Louisville which owns eight of the twelve similar houses doing business in that city and which is known as the largest warehouse company in the tobacco business, handling millions of dollars worth in a year.</p>	Nested tags are implemented with multiple tag refs.

5. Role tagging

Role tagging allows to describe the people involved in the content creation: author, editor, etc. This can be done with an open vocabulary or a dedicated vocabulary, like the MARC Code List for Relators.

■ XML schema

ALTO tag element: <RoleTag>

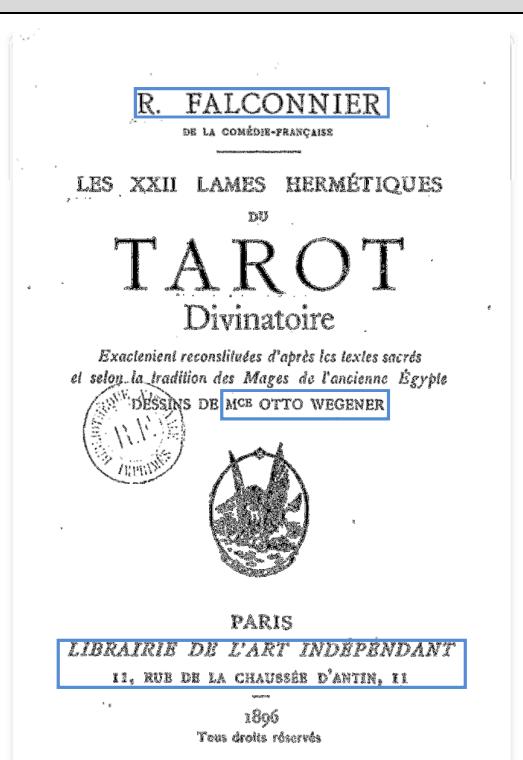
Tag attributes:

- **ID**: Tag id
- **LABEL**: tag type or category
- **TYPE** (optional): Supertype or category of the type
- **DESCRIPTION** (optional): Description of the tag, description of the tag content, or the content itself
- **XmlData** (optional): Any XML encoded metadata helping to describe the tag or the vocabulary used to describe the tag.
- **URI** (optional): Any authority URI relevant for the tag element.

■ Good practices, limitations

- A specific name tagged with a Role tag can also be identified with a NE tag.

Text tags, to be applied on <ComposedBlock>, <TextBlock>, <TextLine> or <String> elements

Use Cases	ALTO tag	Examples
Title pages	<pre><RoleTag LABEL="Author" DESCRIPTION="book's author"/></pre> <pre><RoleTag LABEL="Illustrator" DESCRIPTION="drawings creator"/></pre> <pre><RoleTag LABEL="Publisher"/></pre> <-- Description of the vocabulary used --> <pre><RoleTag LABEL="ill" DESCRIPTION="drawings creator"> <XmlData> <Scheme>marc:relators</Scheme> </XmlData> </RoleTag></pre>	

Foreword, etc.	<pre><RoleTag LABEL="aui" DESCRIPTION="author of the introduction"> <XmlData> <Scheme>marc:relators</Scheme> </XmlData> </RoleTag></pre>	<p style="text-align: center;">II LE TAROT</p> <p><i>Tout le monde, grâce à vous, pourra désormais comprendre, admirer et au besoin consulter ces arcanes mystérieux de la Haute-Science des Mages, car vos Tarots sont d'une lucidité psychique incomparable et leurs maximes si consolantes à tous égards, peuvent également servir à nous diriger dans telle ou telle voie favorable à notre réussite dans la bataille que nous livrons chaque jour à l'inconnu.</i></p> <p><i>Tous mes compliments de sincère admiration pour le savoir et la persévérance que vous avez apportés dans vos recherches et croyez-moi bien, cher Monsieur, cordialement à vous.</i></p> <p style="text-align: right;">A. DE THÈBES.</p>
----------------	--	---

6. Other tags

Any other kind of tags can be described.

XML schema

ALTO tag element: <OtherTag>

Tag attributes:

- **ID**: Tag id
- **LABEL**: Tag type or category
- **TYPE** (optional): Supertype or category of the type
- **DESCRIPTION** (optional): Description of the tag, description of the tag content, or the content itself
- **XmlData** (optional): Any XML encoded metadata helping to describe the tag or the vocabulary used to describe the tag
- **URI** (optional): Any authority URI relevant for the tag element.

7. Reference documents

EPUB 3 Structural Semantics Vocabulary: <http://www.idpf.org/epub/vocab/structure/>

PAGE format: <http://schema.primaresearch.org/PAGE>

MARC Code List for Relators: <http://id.loc.gov/vocabulary/relators.html>

Europeana Newspapers:

- *METS ALTO Profile (ENMAP)*, version 1.0, January 8th 2014
- *Options for the representation of recognized named entities*, version 0.1, December 12th, 2012