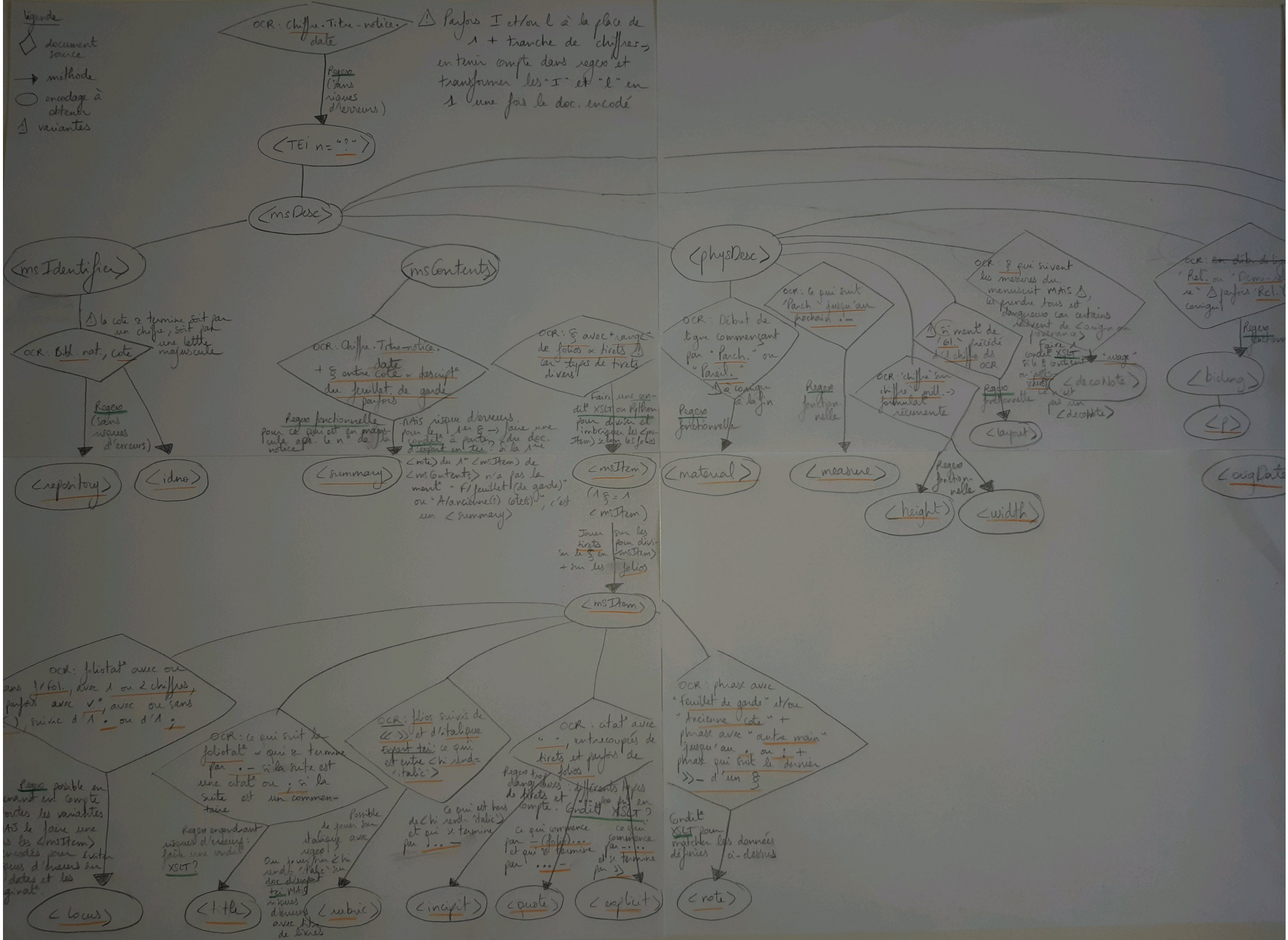
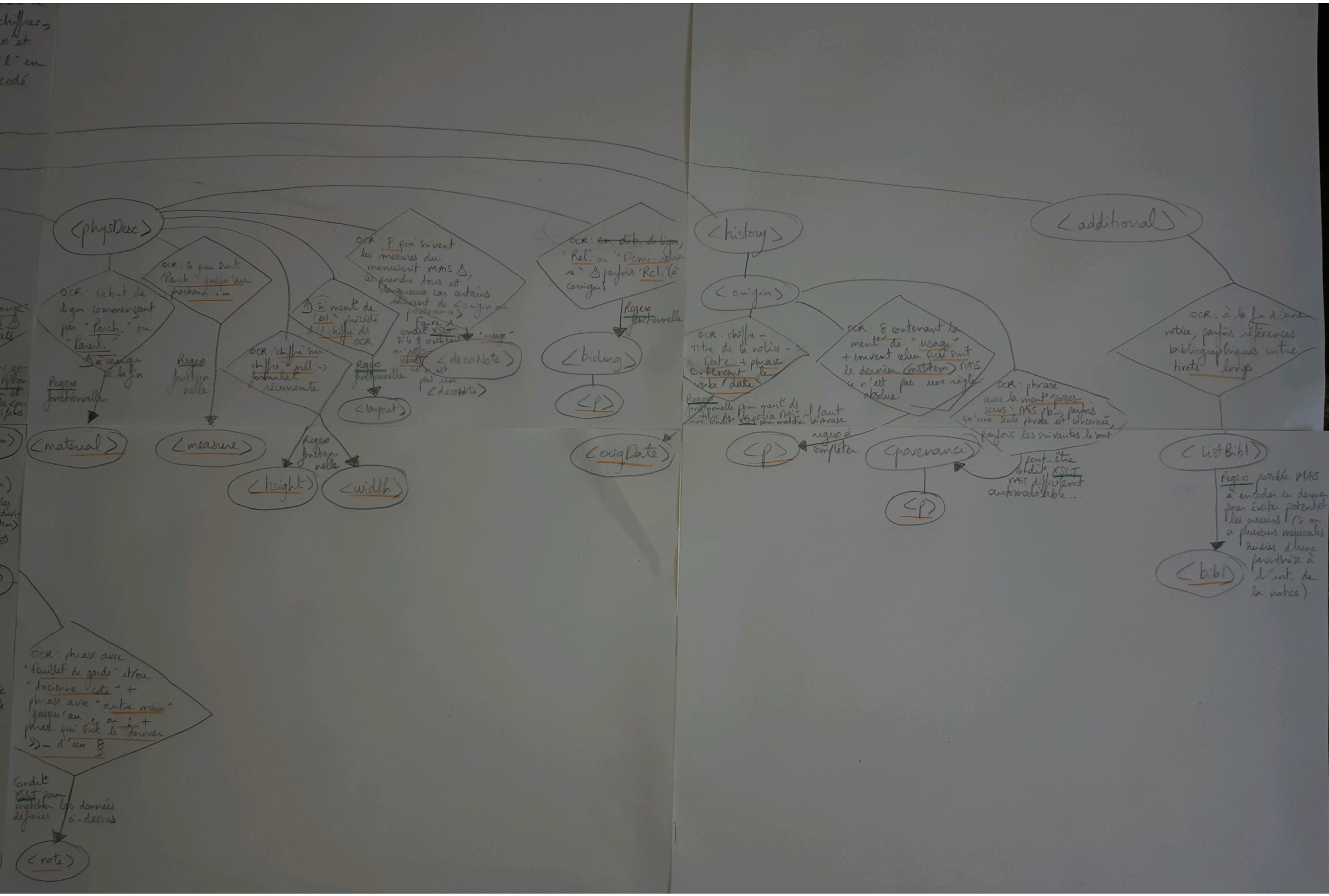

légende
document source
méthode
encodage à obtenir
variantes
OCR : chiffre. Titre-notice. date
Regex (sans risques d'erreurs)
Parfois I et/ou l à la place de 1 + tranche de chiffres, en tenir compte dans regex et transformer les "I" et "l" en 1 une fois le doc. encodé
<TEI n="?">
<msDesc>
<msIdentifier>
la cote se termine soit par un chiffre, soit par une lettre majuscule
OCR : Bibl. nat., cote
Regex (sans risques d'erreurs)
<repository>
<idno>
<msContents>
OCR : chiffre. Titre-notice. + § entre cote & descript° du feuillet de garde parfois
Regex fonctionnelle pour ce qui est en moins après le n° de la notice
MAIS risque d'erreurs pour le 1er § → faire une condit° à partir du doc. d'export en tei : si la 1re <note> du 1er <msItem> de <msContents> n'a pas la mention "F/feuillet (de garde)" ou "A/ancienne(s) cote(s)", c'est un <summary>
<summary>
OCR : § avec "range" de folios & tirets ⚠ ou types de tirets divers
Faire une condit° XSLT ou Python pour diviser et imbriquer les <msItem> selon les folios
<msItem>
(1 § = 1 <msItem>)
Jouer sur les tirets sur le § en pour divi- <msItem> + sur les folios
<msItem>
OCR : folios° avec ou sans f/fol., avec 1 ou 2 chiffres, parfois avec v°, avec ou sans (), suivie d'1 . ou d'1 ;
Regex possible en tenant en compte toutes les variantes MAIS le faire une fois les <msItem> encodés pour éviter plus d'erreurs sur les dates et les foliat°.
<locus>
OCR : ce qui suit la foliotat° & qui se termine par . – si la suite est une otat° ou ; si la suite est un commentaire
Regex engendrant risques d'erreurs : faire une condit° XSLT ?
<title>
Possible de jouer sur italique avec regex ! Ou jouer sur <hi rend="italic"> dans doc. export tei MAIS risques d'erreurs avec titres de livres
<rubric>
OCR : folios suivis de « » et d'italique
Export tei : ce qui est entre <hi rend="italic">
Ce qui est hors de <hi rend="italic">
<incipit>
OCR : otat° avec " . ", entrecoupées de tirets et parfois de folios
Regex trop dangereuse : différents types de tirets et ... pas qu'en compte. Condit° XSLT ?
ce qui commence par – (folio)... et qui se termine par ... –
<quote>
ce qui commence par «... et se termine par »
<explicit>
OCR : phrase avec "feuillet de garde" et/ou "ancienne cote" + phrase avec "autre main" jusqu'au . ou ; + phrase qui suit le dernier »– d'un §
Condit° XSLT pour matcher les données définies ci-dessus
<note>
<physDesc>
OCR : Début de ligne commençant par "Parch." ou "Pareil."
⚠ à corriger à la fin
Regex fonctionnelle
<material>
OCR : ce qui suit "Parch." jusqu'au prochain .–
Regex fonctionnelle
<measure>
⚠ si mention de "cl." précédée d'1 chiffre ds OCR
OCR : chiffre sur chiffre mill. → formulat° récemment
Regex fonctionnelle
<height>
<width>
OCR : § qui suivent les mesures du manuscrit MAIS ⚠, ne prendre tous et dangereux car certains relèvent de <origin> ou <provenance>
Faire ⚠ condit° XSLT si le § contient "wage" ou pas (MAIS ce n'est pas un <decoNote>)
<layout>
<decoNote>
OCR : ~~en déb. de lig~~ "Rel." ou "Demi-re" ⚠ parfois "Rel." conjugué
Regex fonctionnelle
<binding>
<p>
<origPlace>


<physDesc>
OCR : Début de ligne commençant par "Parch." ou "Papier."
⚠ à corriger à la fin
Regex fonctionnelle
<material>
OCR : 6 qui suit "Parch." jusqu'au prochain .–
Regex fonctionnelle
<measure>
OCR : chiffre sur chiffre mill. -> formulat. récurrente
Regex fonctionnelle
<height>
<width>
⚠ Si ment de col. précédé d'1 chiffre ds OCR
Regex fonctionnelle
<layout>
OCR : § qui suivent les mesures du manuscrit MAIS ⚠, à prendre tous et vérifier car certains relèvent de <origin> ou <provenance>
Faire 1 conduit XSLT si le § contient "usage" ou "après" sinon ce n'est pas un <decoNote>
<decoNote>
OCR : ~~en dib. de ligne~~, "Rel." ou "Demi-reliure" ⚠ parfois "Rel." (à corriger)
Regex fonctionnelle
<binding>
<p>
OCR : phrase avec "feuillet de garde" et/ou "ancienne cote" + phrase avec "autre main" jusqu'au . ou : + phrase qui suit le dernier >>– d'un §
Conduit XSLT pour matcher les données définies ci-dessus
<note>
<history>
<origin>
OCR : chiffre – titre de la notice – Date + phrase contenant le verbe / date
Regex fonctionnelle pour mettre ds titre de la notice MAIS il faut un conduit XSLT pour matcher la phrase
regex à compléter
<p>
OCR : § contenant la ment. de "usage" + souvent celui qui suit le dernier <origin> MAIS ce n'est pas une règle absolue
<provenance>
<p>
OCR : phrase avec la ment. "possesseurs" MAIS pb -> parfois qu'une seule phrase est concernée, parfois les suivantes le sont
peut-être conduit XSLT mais difficilement automatisable..
<origDate>
<additional>
OCR : à la fin d'une notice, parfois références bibliographiques entre tirets longs
<listBibl>
Regex possible MAIS à encoder en demi pour éviter potentiels erreurs (si on a plusieurs manuscrits suivis d'une parenthèse à l'int. de la notice)
<bibl>