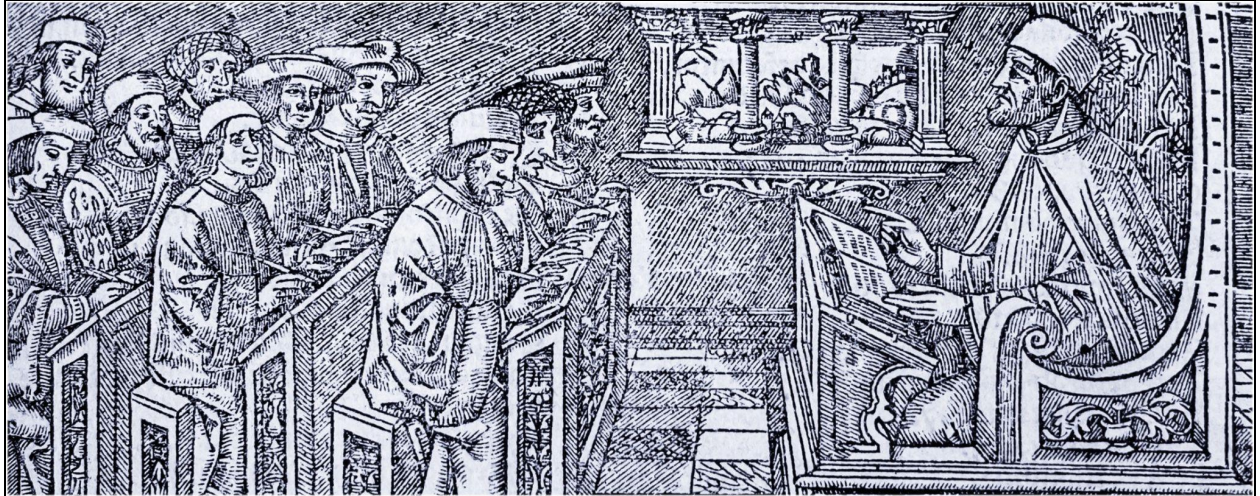


BACK TO ANALOG

THE ADDED VALUE OF PRINTING DIGITAL EDITIONS



Introduction

The topic of TEI 2022 conference and members' meeting — “text as data” — addressed a growing amount and diversity of textual data produced by humanities projects. With the increase of data there is also an expanding need for its quality assurance. Several research data projects have assigned specific teams to tackle the task of standardizing the continuous quality management. I refer, for example, to the [task area “Standards, Data Quality and Curation”](#) within the [NFDI4Culture](#) consortium, or the [KONDA](#) project at the Göttingen State and University Library. The XML production is in fact a process of a continuous validation, correction, and improvement, involving, inter alia, the ODD, RelaxNG, and XML Schemata; custom Python and R scripts; the XSLT, XQuery and Schematron routines integrated into a test-driven development frameworks such as XSpec.

I would like to address a rather unconventional way of testing the TEI XML data, namely printing it. TEI frameworks frequently presuppose HTML and PDF export (e.g. TEI Publisher), the issue I will focus on is the diagnostic value of such prints for the quality control.

Introduction

The School of Salamanca project, its production workflow and printing practice

- Production workflow

- XSL Formatting Objects

- Book components

PDF export as a diagnostic tool

- Order and position of XML elements

 - Marginal notes and milestones

 - Title page and its constituents

- Cross-referencing

- Special characters and abbreviations

- Text mark-up

PDF impact on data quality assurance

The School of Salamanca project, its production workflow and printing practice

The project "[The School of Salamanca. A Digital Collection of Sources and a Dictionary of its Juridical-Political Language](#)" is jointly sponsored by the Academy of Sciences and Literature Mainz, Max Planck Institute for Legal History and Legal Theory and Goethe-University Frankfurt am Main. It aims at creating an online collection of important texts produced by the philosophers, jurists and theologians related to the University of Salamanca — the intellectual center of the Spanish monarchy during the 16th and 17th centuries.

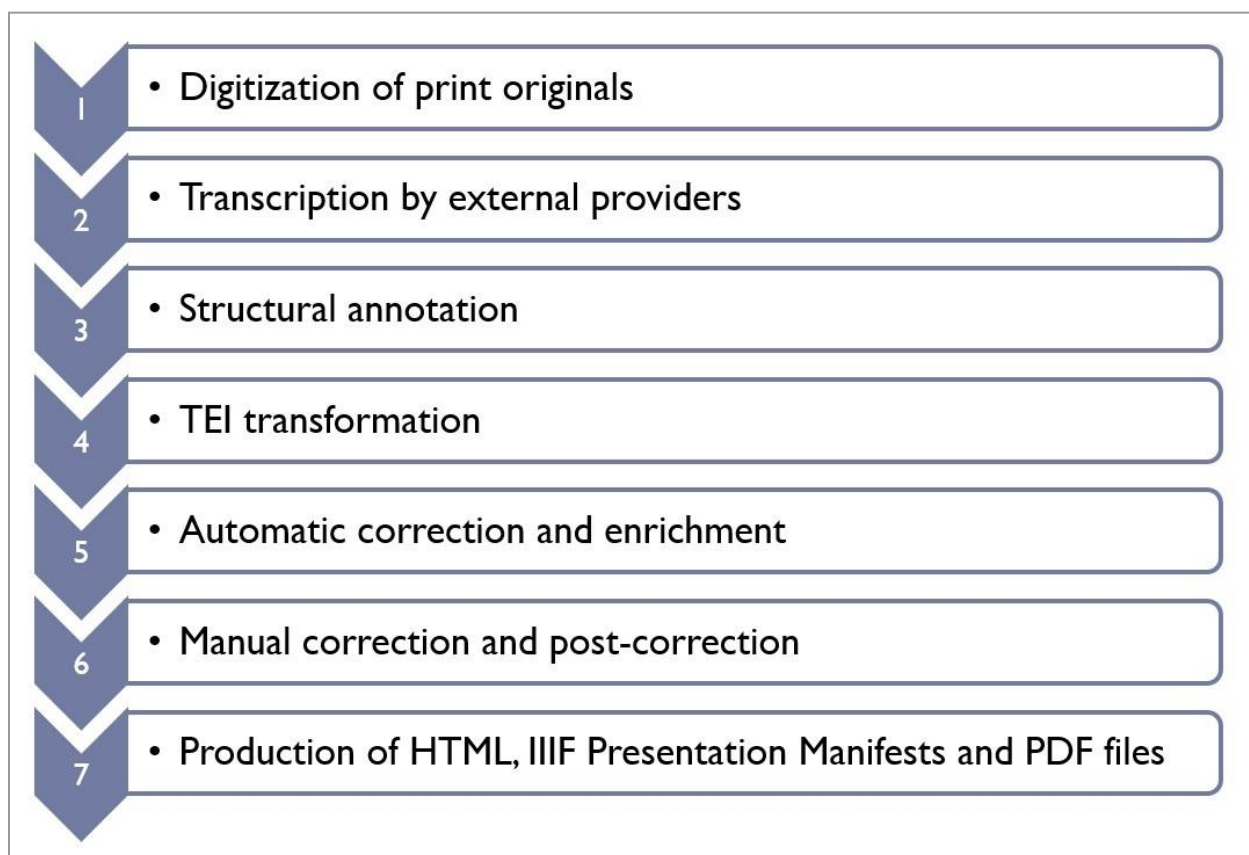
The edition will contain 116 works, with more than 108 000 printed pages of Early Modern Latin and Spanish texts encoded in TEI XML. In addition, we also compose a historic dictionary of approximately 300 essential terms, rendering the importance of the School of Salamanca for the early modern discourse about law, politics, religion, and ethics.

Currently 36 works have completed the production cycle which includes HTML export for online access and full-text search, IIIF Image and Presentation APIs, RDF and TXT export. Recently, the PDF output option was also added, and it had a direct impact on our workflow

and quality control. It is now implemented early in the TEI production as a useful diagnostic tool, exposing semantic and structural inconsistencies of the data.

Production workflow

Salamanca’s production workflow consists of a number of sequential steps, which are familiar to those creating Scholarly Digital Editions (SDEs)¹. It bears similarities to the “waterfall” method, common in software development (Abbas, Gravell, and Wills 2008).



1. The first step is the **digitization** of the print originals held by the libraries worldwide. Already at this stage the XML header is compiled for each work with bibliographic metadata and the facsimiles are published on the IIIF Server.

¹ More on Salamanca’s production workflow, see the recent blogpost: <https://blog.salamanca.school/de/2022/04/27/the-school-of-salamanca-text-workflow-from-the-early-modern-print-to-tei-all/>

2. It is followed by the **transcription**, conducted in **TEI Tite format** by the external providers. TEI Tite, in contrast to the generic TEI All schema, has a reduced, compact vocabulary.
3. The TEI Tite files upon arrival undergo the manual **structural annotation**, cross-referencing and resolution of unclear marks.
4. Only after that the text is automatically **transformed into a project-specific TEI**. The resulting document contains the metadata in <teiHeader>, and all elements and attributes adapted to TEI-All.
5. TEI transformation is followed by a sequence of **automatic corrections and enrichment** routines, driven by XSLT templates. They apply to the annotation of hyphenated words, abbreviations and special characters resolution and xml:id tagging .
6. Then texts go to the **manual correction** and **post-correction**, where the remaining transcription and typographical errors are resolved.
7. The final step is the production and delivery of the **derived data formats** taking place in the Exist DB. These are HTML pages, IIIF Presentation manifests, Search index and crumb trails, plaintext, RDF and PDF files.

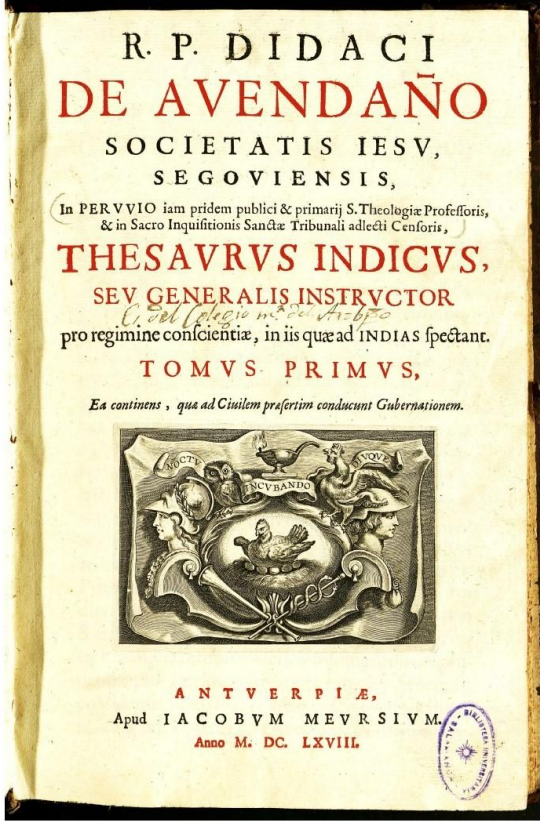
XSL Formatting Objects

Our printing process works through [XSL Formatting Objects \(FO\)](#) technology. Although the latest XSL-FO 1.1 specification dates back to 2006, this format is still widely used. The reasons might lie in the free [Apache FO Processor](#), integrated in Oxygen XML Editor, while its current replacement [CSS Paged Media](#) is proprietary. The latest Apache FOP 2.8 was released in November 2022, and its commercial counterparts, such as [RenderX XEP Engine](#) and [Antenna House Formatter](#), supporting XSL-FO, are regularly updated. At Salamanca we made some preliminary tests with XSL-FO in 2018, and when the production started in 2021, we adopted and upgraded the existing templates. The PDF printout of the released TEI XML files can already be found in our [Github repository](#), on-the-fly PDF generation will be available with the new Salamanca WebApp in 2023.

Book components

In the FO routine the original XML is transformed to XSL-FO document which is then picked up by the Apache FO Processor and converted to PDF. The generic transformation rules apply to all 116 works and render them in print as close to the original as possible. Our template defines nine book components, related to the abstract „page masters” in XSL-FO framework:

- The first component is the so-called “Half title” or “Schmutztitel”, containing the short title of the work and the name of its author. The second is the “Frontispiece”, rendering the scan of the title page of the original.

<p>Thesaurus Indicus, Vol. 1</p> <p>Diego de Avendaño</p>	
---	---

- These are followed by the title page of the digital edition and the edition notice.

<p>Thesaurus Indicus, Vol. 1</p> <p>Diego de Avendaño Antverpiae 1668 (Jacobus Meursius)</p> <p>The School of Salamanca A Digital Collection of Sources and a Dictionary of its Juridical-Political Language https://www.salamanca.school</p> <p>Volume 11.1</p> <p>Directors: Duve, Thomas Lutz-Bachmann, Matthias</p> <p>Akademie der Wissenschaften und der Literatur Mainz Max-Planck-Institut für Rechtsgeschichte und Rechtstheorie Goethe-Universität Frankfurt/Main</p> <p>Electronic publication, 2019-09-25 Online: https://id.salamanca.school/texts/W0001:vol1</p>	<p>Editors:</p> <p>Egjo García, José Luis Rico Carmona, Cindy Glück, David Caesar, Ingo Thönes, Martin Wagner, Andreas</p> <p>Digitized original(s):</p> <p>Bibliotecas de la Universidad de Salamanca https://brumario.usal.es/permalink/34BUC_USAL/1r2q74/alma991005145679705773</p> <p>Proposed citation:</p> <p>Avendaño, Thesaurus Indicus, Vol. 1 (2019 [1668]), in: The School of Salamanca. A Digital Collection of Sources https://id.salamanca.school/texts/W0001:vol1</p> <p><small>This pdf edition does not render the layout of the original print. For detailed information about editorial interventions consult our Edition Guidelines: https://www.salamanca.school/en/guidelines.html. Marginal notes of the original appear as consecutively numbered end notes.</small></p>
---	---

- After that the book contents per se are delivered. First comes the title page of the original rendered from <tei:titlePage>.

Ⓜ

R. P. DIDACI DE AVENDAÑO
SOCIETATIS IESV,
SEGOVIENSIS, In PERVVIO
iam pridem publici & primarij
S. Theologiæ Professoris, &
in Sacro Inquisitionis Sanctæ
Tribunali adlecti Censoris,
THESAVRVS INDICVS, SEV
GENERALIS INSTRVCTOR
pro regimine conscientiæ, in
iis quæ ad indias spectant.
TOMVS PRIMVS, *Ea continens,
quæ ad Ciuilem præsertim
conducunt Gubernationem.*

ANTVERPÆ, Apud IACOBVM
MEVRSIVM. Anno M. DC. LXVIII.
1668

- The introduction section comes next, generated from the <tei:div> elements of <tei:front>, followed by the contents of <tei:body>

<p style="text-align: center;">Diego de Avendaño</p> <p>secūs faxit, confiscatione librorum & aliā graui poenā multabitur, vti latius patet in litteris datis Bruxelle xvi Decembris anni m. dc. lxxvii.</p> <p style="text-align: right;"><i>Signat</i> LOYENS.</p> <p style="text-align: right;">23</p>	<p style="text-align: center;">Thesaurus Indicus, Vol. 1</p> <p style="text-align: center;">1</p> <p>[1] THESAVRI, ET INSTRVCTORIS INDICI TITVLVS PRIMVS.</p> <p style="text-align: center;">De Indiarum jure, & Catholicorum Regum obligationibus circa illarum administrationem.</p> <p style="text-align: center;"><i>Cūm prædictum Ius ampliſſimæ hujus disputationis fundamentum sit, ab illo exordium est optimo iure captandum: Vnde pro eo rogat</i></p> <p>CAPVT I.</p> <p style="text-align: center;">An liceat de jure Indiarum in Catholicis Regibus dubitare.</p> <p>§. I.</p> <p style="text-align: center;"><i>Negatiua Positio stabilitur.</i></p> <p>Stat pro illa id, quod¹⁴ Cardinalis Bellar¹⁵minus <i>Tractatu de Potestate Pontificis in temporalibus</i>, sapienter ostendit, præsertim¹⁶ Cap. 3. vbi benè admonet contra Gulielmum Barclaium, potestatem Pontificis in temporalibus non esse rem dubiam, neque ab opinionibus pendere Doctorum. Id enim sacri Canones variis in locis determinant: pro quo etiam¹⁷ Diana strenuè & eruditè decertat <i>Parte 1. Tract. 2. Resolut. 122. & 123. Parte 10. Tract. 6. Resolutione unica & Tract. 17. Ac primò ex additis, Res¹⁸olutionibus præsertim 8. & 9. Vbi plurimos congerit huic dogmati suffragantes. Et vt alia præteream, veritatem istam luculenter probandam, & definiendam assumpsit Bonifacius Octauus in Extrauag. <i>Vnam Sanctam</i>, in qua¹⁹ cūm de duobus gladiis, spiritali scilicet & materiali, quibus instructa Ecclesia, elegantem discursum præmississet, sic concludit: <i>Parrò subese Romano Pontifici, omni humane creature</i>,²⁰ <i>declaramus, dicimus, definimus, & pronuntiamus omnino esse de necessitate salutis. Sic ille: vbi sensus est iuxta præmissa; subesse inquam omnem humanam creaturam, non solum spiritaliter, sed temporaliter: de spiritali enim²¹ dubium non suberat, cūm Christus Divo Petro omnes oues commiserit; pro quo <i>libitum</i> Pontifex sic locutus: <i>Dicente Domino ipsi Petro: Pisce oues meas. Ioan. 21. v. 17. meas inquit²² & generaliter, non singulariter, has vel illas: per quod commississe sibi intelligitur vniuersas. Ita ille: vnde quod specialiter suā definitione constituit, adtemporalem pertinet potestatem: ex quo orta in Rege Franciæ, & regnicolis illa suspicio, quam Clemens Quintus suā Constitu²³tione submouit, declarans per definitionem & declarationem dictam, nullum Regi & regno præjudicium generari; neque amplius Pontifici subiectum, quam ante prædictam declarationem & definitionem. Et merito quidem: nam Bonifacius suā declaratione & definitione, nouum jus non statuit,</i></i></i></p> <p style="text-align: right;">24</p>
---	---

- The eighth section delivers the contents of <tei:back>.

<p style="text-align: center;">Diego de Avendaño</p> <p>præferant, eorum etiam more vestiti, ad quos debet etiam cura se Protectoris extendere si ipsi velint; nam id forte refugient, vt in suis contractibus agere liberius possint, & domicilium pro libitu commutare. Quod autem de Mularis dictum, de liberis tantum procedit: si enim mancipia sint, quia ex captiui nati, idem est quod de Æthiopiis asserendum; quando, etiamsi liberi essent, eorum deberet attendi paupertas; que quia communiter in prædictis occurrit, iuxta eam iudicandum; cum aliis paupertatis ex abjecta origine incre-[41]mentum miseria dispiciatur; licet se solo non sufficiat: sunt namque multi, sicut & Mestizij adeo lautè viuentes, & in honore habit, vt nullus eos miserabiles reputet, neque in eorum litigijs ad commiserationem peculiarem miserabilibus personis debitam, moueatur. Et quidem viduas, cum diuites sunt, multi inter miserabiles personas numerandas negant, sicut & pupillos, vt videri potest apud Couraruiam <i>Cap. 6.</i> citato. §. <i>Primus casus.</i> Quis enim miserabiles dicat diuites, id habentes, quo miseria possint omnes propulsare? Pecuniam inquam, de qua recens, sed elegans Poëta:</p> <p style="text-align: center;"><i>Me diua omnipotens, cusa, effugiata, rotunda.</i></p> <p>Cujus & Hispana versio illa.</p> <p style="text-align: center;"><i>Aquella, si bien vndida Diosa, que todo lo puede,</i> <i>Redonda, para que rueda, Porque se abata, batida.</i></p> <p>Quod si aliter dicendum, vt apud eundem alij, de quo & <i>suprà num. 154.</i> idè est quia in lege ipsa expressum, in qua de viduis & pupillis mentio, vnde beneficium hoc non debuit eisdem, licet ratio pro ipsis non adeo esset vrgens, denegari. In casu autem nostro nulla est lex, que predictos exprimat, & alioqui ratio non suffragatur ipsis: ex quo fit nullum futurum iudicium, qui vti priuilegijs fauendum in litigijs arbitretur.</p> <p>[sect. 185] Quod si forte aliquis sit, per me certè licebit, & tueri se poterit ex eo quod tales Pontificis priuilegijs Indorum ac Æthiopum in ordine ad res spirituales, seu negotia animæ perfruantur, de quo dictum <i>Titulo 5. numer. 134.</i> quamquam id non adeo solidum sit: pro extensione enim præfata stat Pontificum expressa in quibusdam declaratio, vnde ad alia fit verosimilis ampliatio. Quod secus in his accidit, de quibus agimus: nullus enim legis, aut peculiaris rescripti pro eo indultum reperitur. Sed quia simile extat aliquid, vt in quibusdam alijs accidit, poterit tolerari, & pupillarum ac viduarum diuitum exemplum adscisci in fauorem prædictorum.</p> <p>§. VII. <i>An ex præfatis origo puritatis sanguinis obstet ad officia aliqua, & honorarios titulos requisita.</i></p> <p>[sect. 186] DE hoc Dom. Solorzans <i>citato Cap. 27. numer. 77. & seqq.</i> & in <i>Politica Cap. 29. pag. 242. Col. 1. & seqq.</i> vbi in eorum fauorem resoluit admittendos ad omnia, & ciuilia & sacra, si capacitas suppetat, neque Maurorum aut Iudæorum jure censendos, atque inter eos nobiles ad Equestres Ordines admittendos. Quod ego in Indis rarissimè admitterem: valde enim equestrium Ordinum splendor decresceret & contemptui haberetur, esseque indecorum nimis equestres homines ob ebrietatem ludibrio habitos, quod in talibus potest magno cum fundamento, quisquis eos nouerit, suspicari. Et citatus quidem Auctor de Indorum genere ac Nigrorum promiscue loquitur; cum tamen circa hoc valde diuersus sit in communi æstimatione conceptus: ex Nigris enim oriundi licet ex eo capite possint puritatem habere sanguinis, quia neque Iudæi, neque Mauri: non tamen nobilitatem, quia</p> <p style="text-align: right;">707</p>	<p style="text-align: center;">Thesaurus Indicus, Vol. 1</p> <p>ex mancipijs descendunt, quorum ante captiuitatem genus probari nobile nullo modo potest, vnde & ceteris paribus Indigenis posthabentur.</p> <p>[sect. 187] Quia verò in Societate nostra statutum de puritate dicta extat in <i>Congregat. 5. Decreto 52. & 53. & in 6. Decreto 28.</i> obseruare iuuat eos, qui ex Indorum genere nobiles vulgo habentur, nullum habere inde impedimentum. Sic <i>citato Decreto 28.</i> secus si ignobiles, vsque ad quantum gradum inclusiuè, ex eodem de Tarteris, & alijs ita concludente: intra quos tamen gradus dispensare R. P. Generalem posse, ibidem habetur. Quod autem in eo Decreto Indica sit progenies inclusa, ex eo constat, quia minimè exclusa comperitur. Vbi enim? & quia quando opus est, ad eundem R. P. habitus inuenitur pro dispensatione recursus. Etsi quandoque annuit, nomine dispensationis vtitur: pro quo eorum epistolas poterunt consulere Prouinciales. Si factum aliter aliquando, id certè bonâ fide; quam non suffragari, ibidem declaratum. Id quod in Æthiopijs certius multò.</p> <p>[[42]] INDEX ADDITIONVM AD TOMVM PRIOREM THESAVRI INDICI. Numeri marginales hic tantum indicantur.</p> <p>A</p> <p>Æthiopes. <i>Gaudent Indorum priuilegijs quoad temporalia, quia miserabiles persone. 182. & 183</i> <i>Ab ipsis origo an obstet sanguinis puritati ad multa requisita. 186</i></p> <p>Alexander VI. <i>Illius defensio, & singularis deuotio ergo Eucharistiam. 2</i></p> <p>Angelus Maria Verricelli. <i>Illius ergo Indiarum expugnationem & donationem Pontificiam asserta excutuntur, cum donationis eiusdem defensione. 3. & seqq.</i></p> <p>B</p> <p>Bellum. <i>AN inferri possit nolentibus Euangelij Predicatore admittere. 8</i></p> <p>Beneficium. <i>Circa Ecclesiastica, vide v. Electio.</i></p> <p style="text-align: right;">708</p>
---	---

- The publication concludes with endnotes.

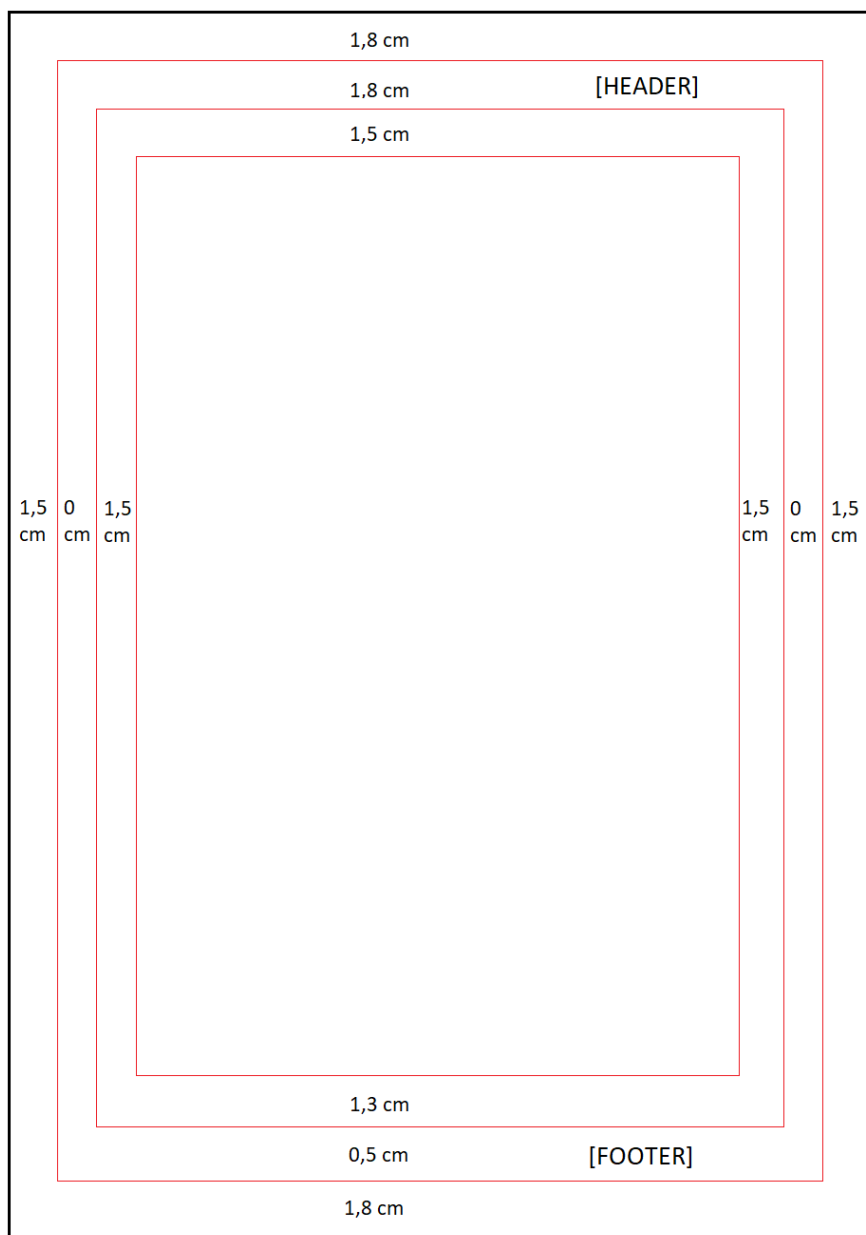
<p>Diego de Avendaño</p> <p>NOTAE</p> <ol style="list-style-type: none"> 1 Sic Ordinatione, de quibus Titul. 5. Cap. 9. 2 Eccl. 1. r. 21. 3 Cyprianus Lib. 3. 33. 4 Prou. 8. r. 12. 5 Sapient. 10. r. 4. 6 Horatii Epist. 1. Lib. 2. 7 Goffridus. 8 Inamali. 9 Tomo 1. Problemat. in Prefat. 10 2. Cor. 4. v. 7. 11 Prou. 2. r. 4. 12 Gennadius. 13 Origenes. 14 N. 1 15 Card. Bell. 16 Papam habere potestatem in temporalibus, dogma certum. 17 Diana. 18 Pontificia pro eo definitio. 19 Duo gladij à Christo dati. 20 Extrat. Vnam Sanctam. 21 Omnes oves commisit. 22 Ioan. 21. v. 17. 23 Ex Bonifacij VIII. Constitutione non est noua inducta subiectio. 24 Nullam ius declaratio inducti, neque alteri periclitat. 25 Marriot. 26 Pontis superior ratione monstratur. 27 Enasio preclusa de soli in Catholicis Regibus protectione. 28 Dominium datum ostenditur. 29 Et inuestituræ. 30 Quid illa vide P. Olate disp 117. de Contractib. Cap. Per tuas. 31 Fraus non est in Ecclesia actibus. 32 Donatio Regibus Portugallie à Sede Apostolicâ facta. 33 Sine illâ nullam in Christianis Principibus dominium, in quo deceptus Cioquens. 34 Alie Pontificum concessiones. 35 Alexandri VI. probitas propugnata. <p style="text-align: right;">717</p>	<p>Thesaurus Indicus, Vol. 1</p> <ol style="list-style-type: none"> 36 Opusculi. Bazis. Soloran. 37 Card. Bellarminus pro Pontificia donatione militat. 38 Errasse Alexandrum aliquorum assertio. Gryphus. 39 Valide confutata, & notata. 40 Pontifex, esse dominium civile non habet, transferre illud ab vno in alium potest. 41 Quisnam alij id possint. 42 Diana. 43 Heretici donationi infens Pontifici. 44 Philippi Mornei blasphemia. 45 Manifesti depulsa. 46 Errasse in facto, & non in iure. Pontificem. aliquorum assertio. 47 P. Valentia. 48 P. Bezan. 49 Confutatur illi. 50 Factum à iure quomodo nequeat separari. 51 Probabilem sententiam scutum Pontificem, aliorum pronuntiatum. 52 efficaciter improbatum. 53 Ioannes Zapata. 54 Ecclesie potestas ad propagandam fidem certum donationis fundamentum. 55 D. Thom. 56 Ex P. Suario luculenta probatio, D. Thome doctrinam amplecti. 57 Non requiritur experientia impedimentorum, sed periculum sufficit, contra Durandum. 58 Subditi, & serui ea de causa possunt ex dominorum excimi potestate. 59 D. Thom. 60 Ad periculum subuersionis ratio D Thome generalior reuocatur. 61 Fideles infidelibus ut seruire quant. 62 Etiam infideles principes conuertantur, sua adducta doctrina. 63 P. Saar. 64 Christiani Principes circa supernaturalia aliquo iurisdictionem exercent. 65 Sola ad regimen conuersorum idonei. 66 Dom. Soloranus. 67 Sententia Pape sti legitime recipiende. 68 Cap. Cum inter vos. 69 Cap. Absit. 70 Et sti D. Petri oracula. 71 Cap. Sic omnes. 72 Nicolaus Pape. <p style="text-align: right;">718</p>
--	---

This generic template allows duplicating and removing the components dynamically if, for example, the original contains two title pages (one of the volume series and one of the current volume), or when it lacks the introduction section and the pagination should start from the body.

The canons of Western book page design place the center of the text area above the center of the page and define the gutter margin narrower than the fore-edge margin². As our PDF publication is supposed to be printed or viewed as A4 pages (height="29,7 cm" width="21,0 cm"), we simplified this layout: the distance to the text on the left and right is equal (3,0 cm), on top it measures 5,1 cm and at the bottom - 3,6 cm. The XSL-FO “page masters”

² For the description of this page layout, see, for example, “Introduction to CSS for Paged Media” pp. 21-22, available for download at <https://www.antennahouse.com/css>.

prescribe three types of margins on each page side, two of which are not printable (they remain empty), and another one which contains a header (with the work title and the author) and a footer (with page numbers). Below is a schematic layout of the page with the dimensions of the margins:



We found it quite useful to define the print area in XSL-FO as a one-cell table: its borders can be activated in debugging sessions to control the margins and other layout features.³

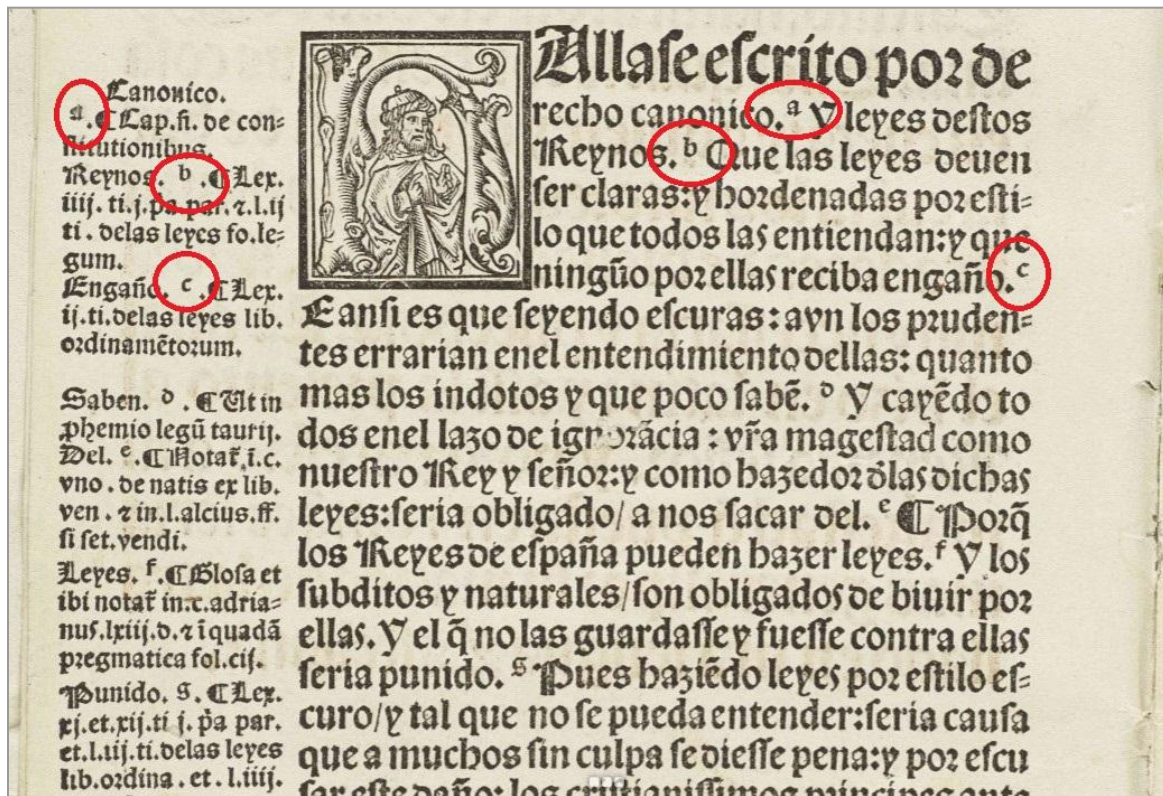
PDF export as a diagnostic tool

PDF production helped us to diagnose some of the XML problems. The XML issues we encountered were of different types and apply to a) the order and the position of elements, b) cross-referencing, c) character encoding, and d) text mark-up.

Order and position of XML elements

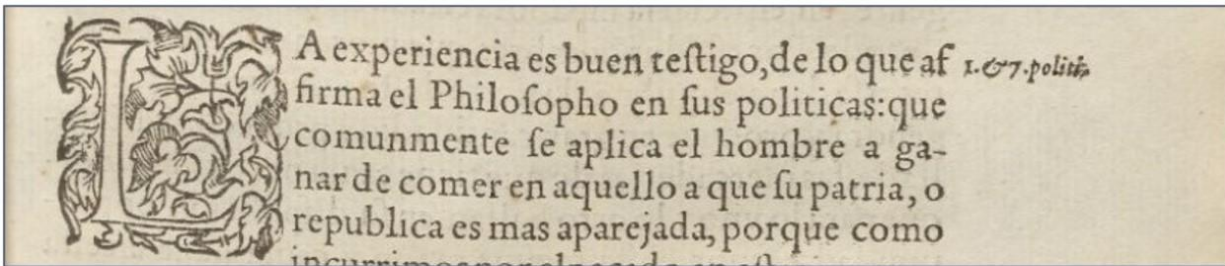
Marginal notes and milestones

Marginal notes at Salamanca usually contain two elements – the anchor in the text and the note body itself. The anchors can be alphabetical, numerical, or expressed with a symbol.




³ A detailed description of the XSL-FO template is available in Salamanca's Github repository: <https://github.com/digicademy/svsal/tree/master/xslt-fo-pdf#readme>.

Sometimes the anchor is missing in the main text, and the note is just located next to the line it refers to. In XML this corresponds to the position before the <lb> element. The number of marginal notes in one document can be up to 8,000.



```
LA experiencia es buen testigo, de lo que af<note place="margin" anchored="false"
xml:lang="la" xml:id="xx">NOTE</note><lb break="no" rendition="#noHyphen"
xml:id="W0007-00-0035-lb-0006"/>firma el Philosopho en sus politicas: que
```

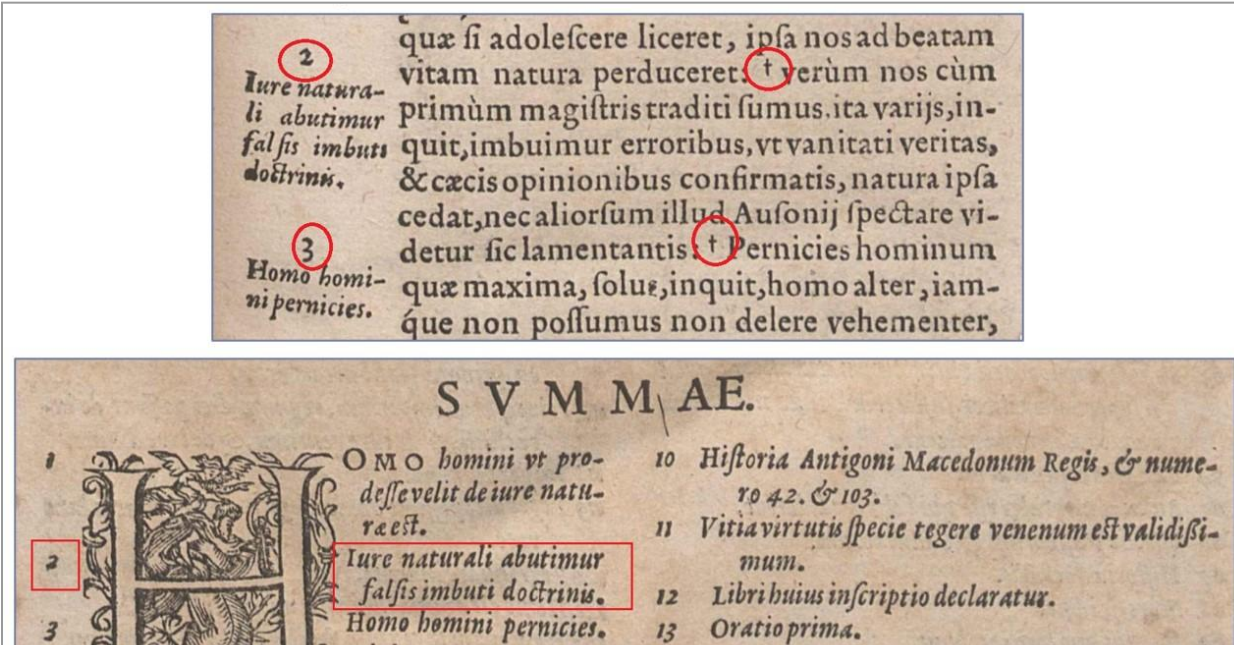
A substantial difficulty of displaying non-anchored marginal notes in PDF is conditioned by the fact that in the original they “float” alongside the main text. In XML they are encoded before the line beginning (<lb>), which does not correspond to the line beginning in PDF. Apache FO processor has insufficient support for floats, and the formatters which fully support them, such as Antenna house or RenderX, are proprietary. Considering this, we decided to unify the representation of anchored and unanchored notes – they all get a numerical mark and an anchor in the main text in PDF. In addition, we place all the notes as endnotes in a separate section, creating cross-references between them and anchors.



Hallase escrito por derecho canonico.¹ y leyes destes Reynos.² Que las leyes deuen ser claras: y hordenadas por estilo que todos las entiendan: y que ninguno por ellas reciba engaño.³

1 Cap. fi. de constitutionibus.
2 Lex. iiii. ti. j. pa. par. et l. ij. ti. delas leyes fo. legum.
3 Engaño. c. 6. l. 1.

Milestones have a structure similar to marginal notes – they have an anchor in text and a related mark in the margin, usually an Arabic number. In addition, they frequently have another component – an item in summary which they are linked to.



Homo homini perniciosa. *quæ si adolescere liceret, ipsa nos ad beatam vitam natura perduceret. † verum nos cum primùm magistris traditi sumus. ita varijs, inquit, imbuimur erroribus, vt vanitati veritas, & cæcis opinionibus confirmatis, natura ipsa cedat, nec aliorum illud Aufonij spectare videtur sic lamentantis. † Pernicies hominum quæ maxima, solue, inquit, homo alter, iamque non possumus non delere vehementer,*

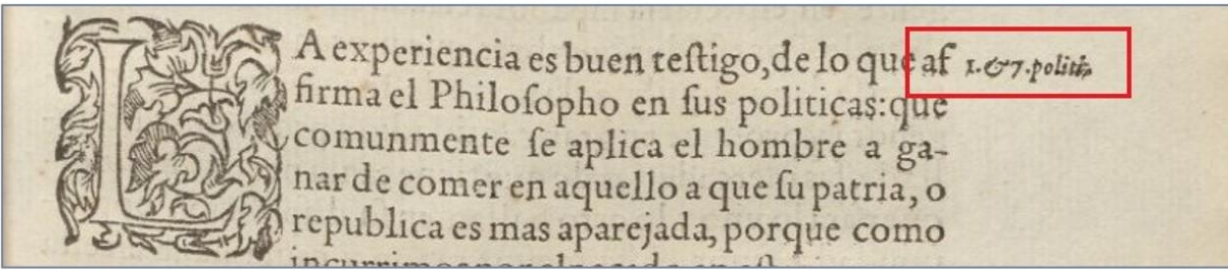
2 Iure naturali abutimur falsis imbuti doctrinis.
3 Homo homini perniciosa.

S V M M A E.

1 *H*omo homini vt prodesse velit de iure natura est.
2 Iure naturali abutimur falsis imbuti doctrinis.
3 Homo homini perniciosa.

10 Historia Antigoni Macedonum Regis, & numero 42. & 103.
11 Virtus virtutis specie tegere venenum est validissimum.
12 Libri huius inscriptio declaratur.
13 Oratio prima.

PDF export highlighted a particular aspect of the encoding of non-anchored marginal notes and milestones, situated at line end containing a word break. In this case the anchor appears in the middle of the word.



Cap. I. Del intento del Autor.

LA experiencia es buen testigo, de lo que **af^dfirma** el Philosopho en sus politicas: que comunmente se aplica el hombre a ganar de comer en aquello a que su patria, o republica es

```
<hi rendition="#initCaps">L</hi>A experiencia es buen testigo, de lo que  
<note place="margin" anchored="false" xml:lang="la" xml:id="xx">NOTE</note>  
af<lb break="no" rendition="#noHyphen" xml:id="V0007-00-0035-lb-0006"/>firma  
el Philosopho en sus politicas: que
```

This case pointed at the importance of data visualization already during the design phase of data modeling. From the XSL-FO perspective marginal notes and milestones should have been encoded in TEI not at word break, but before or after the word they refer to.

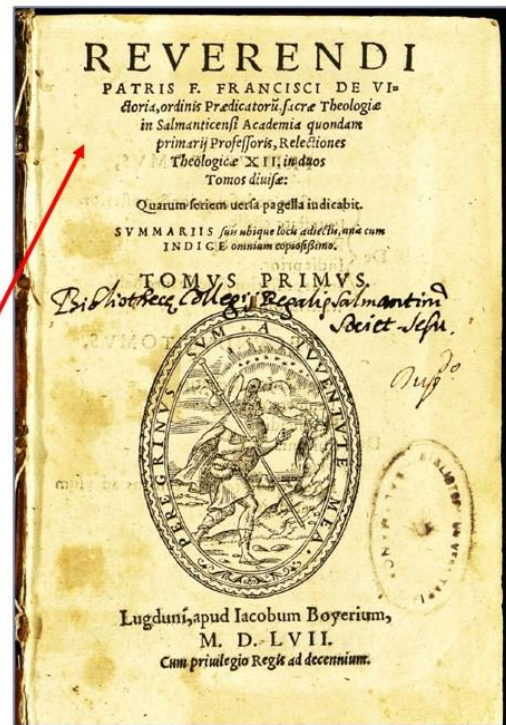
Title page and its constituents

The PDF printout can be used to test rules regarding the order of XML elements. In the title page below the first paragraph was encoded in the element <byline>, which in all other works follows the title. Correspondingly, it was printed in the wrong position.

Relectiones Theologicae
XII. in duos Tomos diuisae:
Quarum seriem uersa pagella
iudicabit. SVMMARIIS suis
ubique locis adiectis, una cum
INDICE omnium copiosissimo.

TOMVS PRIMVS.

[[1]] REVERENDI PATRIS F. FRANCISCI DE
Victoria, ordinis Praedicatorum, sacrae Theologiae in
Salmanticensi Academia
quondam primarij Professoris, Lugduni,
apud Iacobum Boyerium, M. D.
LVII. Cum priuilegio Regis ad decennium.



Alternatively, an item will not be rendered in PDF if it is located in a different position in the hierarchy than prescribed by the template.

Cross-referencing

The PDF layout is a useful tool for checking the cross-references. In a given example the item in the “Summary” is lacking cross-reference, but the reason for that is actually a duplicated milestone number it refers to. As a collateral bug of another improvement, this exceptional behavior was not caught up by Schematron:

Fernando Vázquez de Menchaca

122 Humanum ingenium procliue ad dissentiendum.

123 Magistratus quilibet largè dicitur rex.

124 Reges quare non habeant, aut egeant bruta.

125 Regibus, aut magistratibus subesse, non est necessitatis præcise.

126 Quo maior administratio, eò sublimior administrans.

127 Doctores actu legentes præferuntur ijs qui actu non legunt.

[2r] 128 Dignior est qui digniori loco præest.

129 Filius an sit patri ratione officij præferendus.

130 Historia per Aulum Gellium relata.

131 Ciues an digniores quàm municipales.

132 Hispania digniſſima.

133 Loci consideratio magni momenti est.

134 Historia Alexandri.

135 Dignior est is, cui Princeps magis fauet.

The same effect may occur if the reference is made to a separate document, not included in the current publication. This functionality can be implemented in HTML, but in PDF it leads to a missing cross-reference .

The actual design of a reference is another feature which can be tested – namely, which part of the string should be highlighted and serve as an anchor? In the example below two different conventions were chosen for referencing the section of the book from its summary.

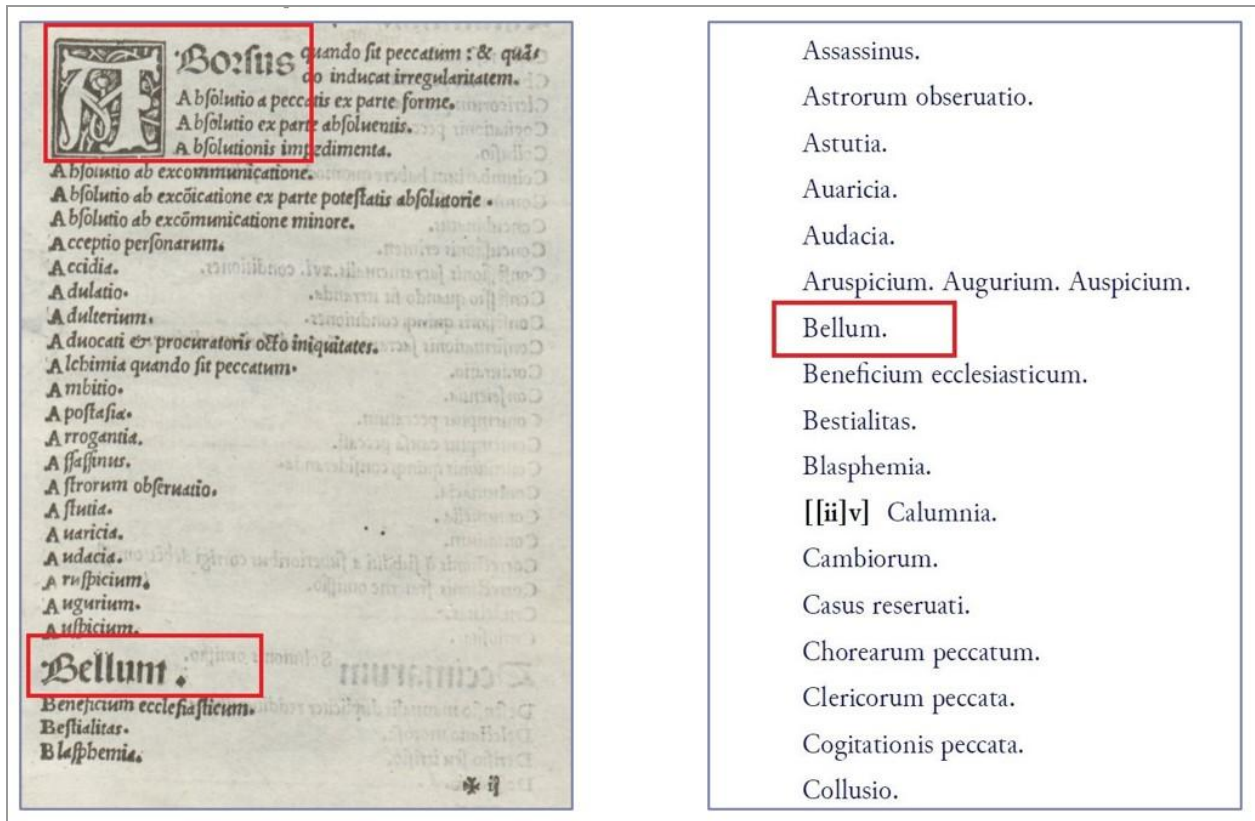
<p>Liber XI. BAPTISMVS.</p> <p>SECTIO I. Receptiores de Baptismo sine pugna sententiæ.</p> <p>Cap. I. DE Nomine, Definitione, Institutione, differentia. ibid</p> <p>II. De Sacramenti Baptismi necessitate. 2</p> <p>III. De Materia Baptismi remota. 4</p> <p>IV. De Materia Baptismi proxima. ibid.</p> <p>V. De Baptismi forma. 5</p> <p>VI. De Baptismi subiecto. 6</p> <p>VII. De dispositione in suscipiente Baptismi. 7</p> <p>VIII. De Baptismi ministro. 9</p> <p>IX. De Baptismi effectibus. 10</p> <p>X. De Baptismi solemnitate. 11</p> <p>XI. De Baptismo sanguinis, seu de Martyrii. 12</p>	<p>Liber XII. CONFIRMATIO.</p> <p>SECTIO I. De Confirmatione receptiores sine pugna sententiæ.</p> <p>89</p> <p>Cap. I. DE Confirmationis intentione. ibid.</p> <p>II. De remota confirmationis materia. ibid.</p> <p>III. De materia Confirmationis proxima. 90</p> <p>IV. De sacramenti Confirmationis forma. 91</p> <p>V. De Confirmationis Effectibus. 92</p> <p>VI. De Confirmationis subiecto. ibid.</p> <p>VII De Sacramenti Confirmationis Ministro. 93</p>
--	---

Special characters and abbreviations

PDF renders a constituted version of Salamanca TEI encoding, meaning that selected special characters and abbreviations are expanded. We display the standardized versions of characters such as Latin 7 (et) and Latin ƒ (long s). In addition, Latin, Greek, Hebrew, and Arabic characters occur and need special fonts to be rendered properly. PDF export thus delivers handy templates for copy-editing and proofreading.

Text mark-up

PDF helps to control the uniformity of text mark-up. This can apply to capitalization, italicized script, superscript, initial, and bold parts of the text.



In the following example two functionally identical passages have different encodings, the second one marked as header .

[[v]] *FACULTAS R. P. PROVINCIALIS Societatis Iesv, in Prouincia Castellana.*

EGo Christophorus de los Cobos Prouincialis Societatis Iesv in Castellana Prouincia. Potestate ad id mihi facta à Reuerendo admodum Patre Claudio Aquaiuia Præposito nostro generali, facultatem facio, vt *Tractatus de Legibus* à P. Ioanne de Salas nostræ Societatis, sacræ Theologiæ professore compositus, & eiusdem Societatis grauium, doctorúmque hominum iudicio approbatus typis mandetur, in quorum fidem has literas manu nostra subscriptas, & sigillo nostro munitas dedimus Vallisolti vigesima quinta die mensis Augusti Anni 1608.

C. DE LOS COBOS.

[[vi]] *FACULTAS R. P. PROVINCIALIS Societatis Iesv, in Prouincia Lugdunensi.*

LVdouicus Michaelis Prouincialis Societatis Iesv, in Prouincia Lugdunensi, iuxta priuilegium eidem Societati à Rege Christianissimo Henrico III. die 10. mensis Maij anni 1583. Et Henrico IV. 20. Decemb. 1606. concessum, quo Bibliopolis omnibus prohibetur, ne libros ab eiusdem Societatis hominibus compositos, absque illorum permissione imprimant; permittit Ioanni de Gabiano Lugdunensi Bibliopolæ, vt *P. Ioannis de Salas Gumielensis è Prouincia Castellana, Societatis Iesv tractatum de Legibus, in secundam secundæ sancti Thomæ: ad sex proximos annos*

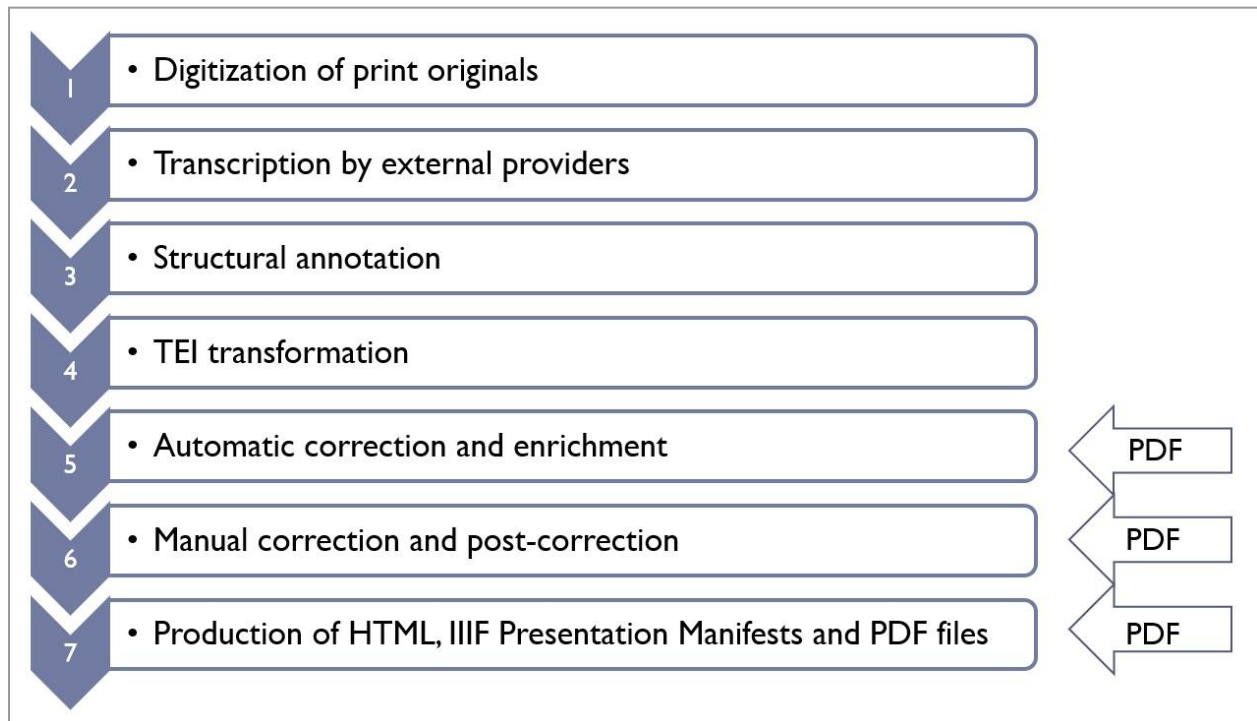
PDF export visualizes the semantics of the encoding – the aspect which cannot be controlled by the XML Schemata and Schematron.

PDF impact on data quality assurance

PDF creation thus assists the data quality assurance in two ways: on the one hand, it raises formal errors in the code and on the other hand — it facilitates manual corrections by providing a new "optical impression" of the data where certain kinds of inconsistencies

stand out more clearly to a human eye. Some of the raised issues depend on the definition of an “error”, which in the context of scholarly digital editions is a rather challenging task. In the software world “bug” is a deviation from the desired behavior defined in the software specification (Crispin 2008: 416). The descriptive nature of XML-based digital editions includes an element of ambiguity and interpretation (Wenz, Kesper, and Taentzer 2022). The identification of an error thus requires a skill set which is different from that of a traditional software tester.

I have mentioned above that at Salamanca the steps of the production pipeline are executed sequentially, in a “waterfall” model, where the output of each stage is validated by the respective Schematron file. The final XML is then controlled by the [RelaxNG schema](#). The PDF generation was initially intended to be one of the export methods of the TEI data, located at the very end of the product development. As soon as it was implemented, we realized that this type of data visualization can be used as a diagnostic tool to expose semantic and structural discrepancies in the source data. We therefore moved PDF export up the TEI production workflow.



This implementation of PDF print bears similarity to the so-called Agile development method. In Agile the software product is built in small progressive chunks, and each of the development cycles includes feature clarification, design, coding, *and testing*. It is conducted by cross-functional teams of people who house a range of expertise including programming, testing, analysis, database administration, user experience and infrastructure (Black 2017: 7).

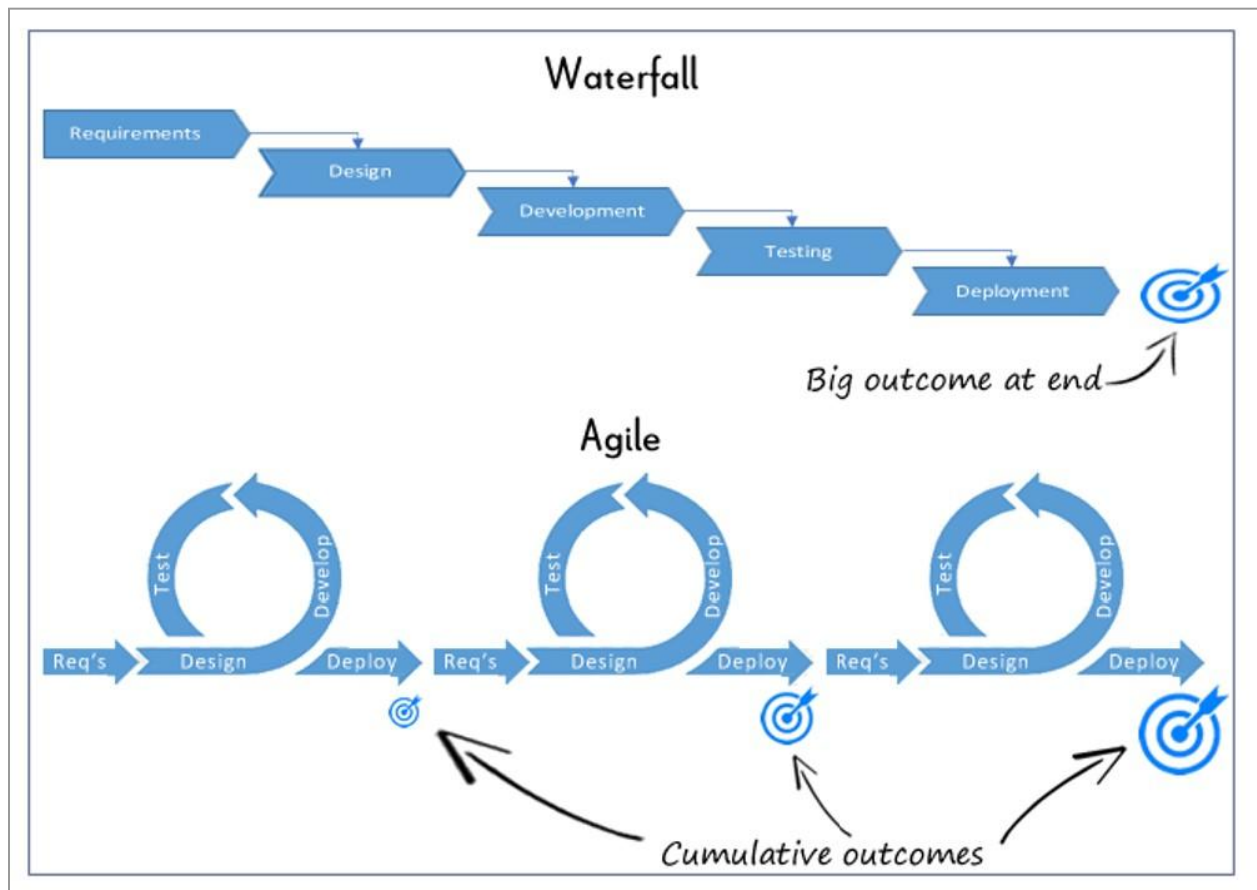


Image source: <http://crmsearch.com/images/agileandwaterfall.gif>

PDF creation at Salamanca abides by the principles of Agile software testing. It does not only start early in the data development and repeats with every subsequent step. It also breaks the traditional boundary between the software developers and researchers: the latter virtually function as testers providing direct feedback to the encoding team and thus actively participating in the development process. Researchers are used to dealing with a

high degree of uncertainty, where not the software specification, but the knowledge of the subject and intuition are necessary to distinguish a “bug” from a “feature”. These two aspects — “fuzziness” of an error and a researcher skill set required from a tester — is what differentiates the conventional software quality assurance from quality assurance in digital humanities (QA4DH).

Maxim N. Kupreyev

The School of Salamanca Project

Bibliography

- Abbas, Noura, Andrew M. Gravell, and Gary B. Wills. 2008. “Historical Roots of Agile Methods: Where Did ‘Agile Thinking’ Come From?” In *Agile Processes in Software Engineering and Extreme Programming*, edited by Pekka Abrahamsson, Richard Baskerville, Kieran Conboy, Brian Fitzgerald, Lorraine Morgan, and Xiaofeng Wang, 94–103. Lecture Notes in Business Information Processing. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-68255-4_10.
- Black, Rex, ed. 2017. *Agile Testing Foundations: An ISTQB Foundation Level Agile Tester Guide*. Swindon, UK: BCS, The Chartered Institute for IT.
- Crispin, Lisa. 2008. *Agile Testing: A Practical Guide for Testers and Agile Teams*. 1st ed. Upper Saddle River, NJ: Addison-Wesley Professional.
- Wenz, Viola, Arno Kesper, and Gabriele Taentzer. 2022. “Classification of Uncertainty in Descriptive Data Representing Scientific Knowledge,” March. <https://doi.org/10.5281/zenodo.6327011>.