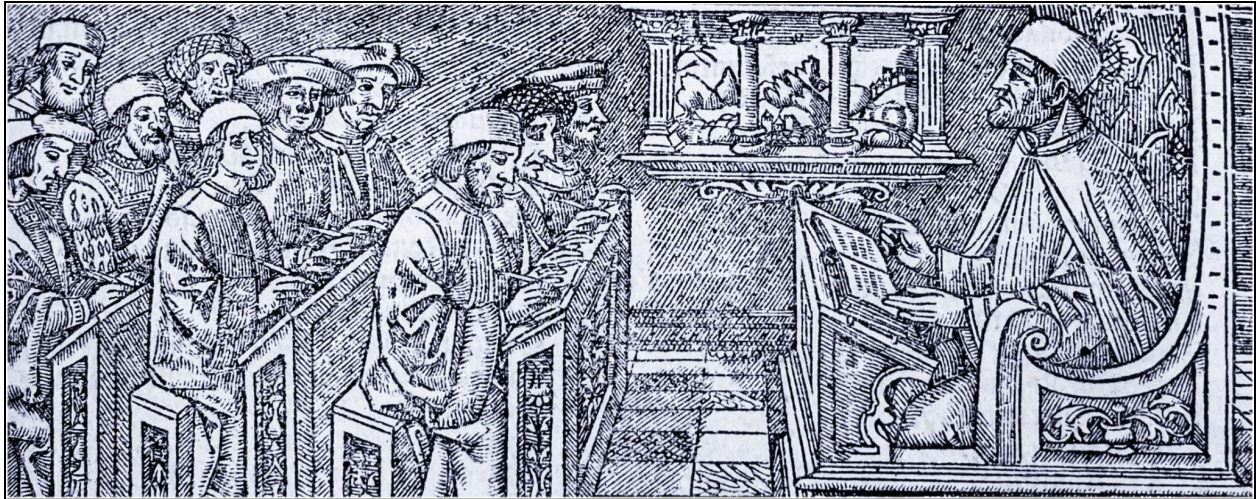# BACK TO ANALOG
THE ADDED VALUE OF PRINTING TEI EDITIONS

Introduction

The topic of this year's TEI conference and members' meeting — "text as data" — addresses 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 even assigned specific teams to tackle the task of standardizing the continuous quality management. I refer, for example, to the German NFDI4Culture consortium which includes the task area "Standards, Data Quality and Curation"[1], or the KONDA project at Bildarchiv Foto Marburg[2]. The XML data production is in fact a process of a continuous validation, correction, and improvement, involving, inter alia, custom Python and R scripts; ODD, RelaxNG, and XML Schema; the XSLT, XQuery and Schematron routines integrated into a test-driven development frameworks such as XSpec.

---

[1] https://nfdi4culture.de/what-we-do/task-areas/task-area-2.html
[2] https://zenodo.org/communities/konda-project

In today's talk I would like to address a rather unconventional way of testing the TEI data, namely printing it. TEI production pipelines frequently presuppose HTML or PDF export (e.g. TEI Publisher), the issue I will focus on today is the diagnostic value of such prints for quality control.

The School of Salamanca project, its production pipeline 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[3]. We aim 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.
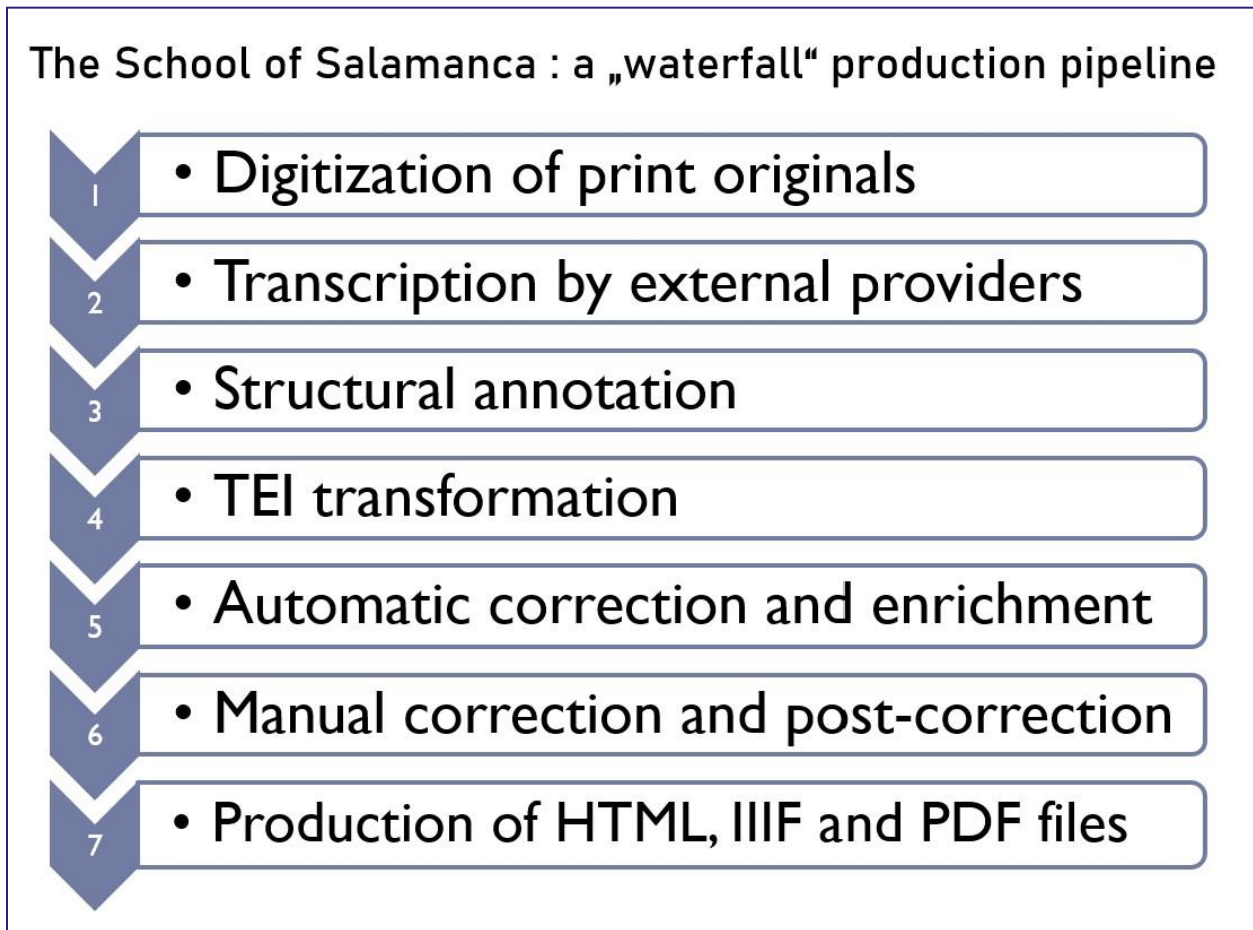
The edition will contain 116 works, including more than 108 000 printed pages of Early Modern Latin and Spanish texts encoded in TEI XML. In addition, we also compose a historic dictionary of approximately 300 essential terms, rendering the fundamental importance of the School of Salamanca for the early 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.

---

[3] https://salamanca.school/

Production pipeline

Our production pipeline consists of a number of steps, which are familiar to everyone creating SDEs (Scholarly Digital Editions)[4].

The School of Salamanca : a „waterfall" production pipeline

1. • Digitization of print originals
2. • Transcription by external providers
3. • Structural annotation
4. • TEI transformation
5. • Automatic correction and enrichment
6. • Manual correction and post-correction
7. • Production of HTML, IIIF and PDF files

1. The first step is the **digitization** of the print originals held by 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.

---

[4] More on Salamanca's production pipeline at:
https://blog.salamanca.school/de/2022/04/27/the-school-of-salamanca-text-workflow-from-the-early-modern-print-to-tei-all/

2. This is followed by the **transcription** by the external providers, conducted in TEI Tite format, which, compared 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 at the fourth step the text is automatically **transformed into** a project-specific **TEI**. The resulting document contains both 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, the 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, plaintext and RDF data creation, and PDF files.

Is XSL-FO dead?

Our printing process works through XSL Formatting Objects (FO) technology. It was proclaimed "dead" around 10 years ago: the latest XSL-FO 1.1 specification dates to 2006[5] and the XSL-FO working group was closed in 2013. A current proposal for its replacement is CSS Paged Media Module Level 3[6]. Surprisingly, the technology is still alive and still "kicking". The reasons for this might lie in a free Apache FO Processor, integrated in Oxygen Author, while its current counterpart CSS Paged Media is proprietary. The latest Apache FOP 2.7 was released in January 2022, and the proprietary Antenna House Formatter v7 supports both XSL-FO and CSS Paged Media. At Salamanca we made some preliminary tests with XSL-FO in 2018, and when we finally started the production in 2021, we adopted and upgraded the existing templates.
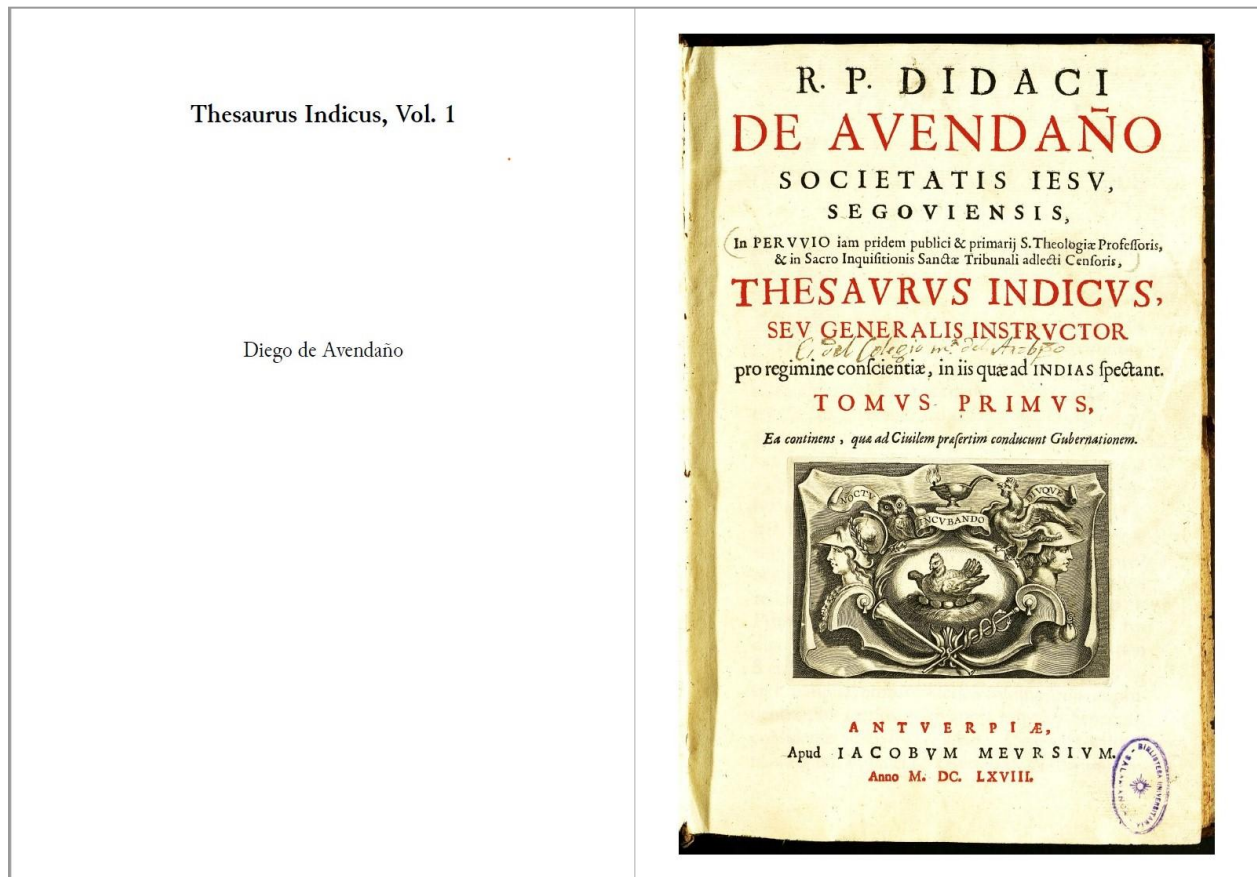
---

[5] https://www.w3.org/TR/xsl11/
[6] https://www.w3.org/TR/css-page-3/

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 most challenging task is to create generic rules, which apply to all 116 works and render them in print as close to the original as possible. Our template defines nine book components, referring in their turn to more abstract „page masters":

- 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.

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

Thesaurus Indicus, Vol. 1

Diego de Avendaño
Antverpiae 1668 (Jacobus Meursius)

The School of Salamanca
A Digital Collection of Sources and a
Dictionary of its Juridical-Political Language
https://www.salamanca.school

Volume 11.1

Directors:
Duve, Thomas
Lutz-Bachmann, Matthias

Akademie der Wissenschaften und der Literatur | Mainz
Max-Planck-Institut für Rechtsgeschichte und Rechtstheorie
Goethe-Universität Frankfurt/Main

Electronic publication, 2019-09-25
Online:
https://id.salamanca.school/texts/W0001:vol1

Editors:

Egio García, José Luis
Rico Carmona, Cindy
Glück, David
Caesar, Ingo
Thönes, Martin
Wagner, Andreas

Digitized original(s):

Bibliotecas de la Universidad de Salamanca
https://brumario.usal.es/permalink/34BUC_USAL/1r2qv74/
alma991005145679705773

Proposed citation:

Avendaño, Thesaurus Indicus, Vol. 1 (2019 [1668]), in:
The School of Salamanca. A Digital Collection of Sources
https://id.salamanca.school/texts/W0001:vol1

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.

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

[[i]]

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.*

ANTVERPIÆ, 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>

Diego de Avendaño

secùs faxit, confiscatione librorum & aliâ graui pœnâ mulctabitur, vti latiùs patet in litteris datis Bruxellæ xvi Decembris anni m. dc. lxvii.

Signat LOYENS.

Thesaurus Indicus, Vol. 1

1

[1] THESAVRI,   ET INSTRVCTORIS INDICI   TITVLVS PRIMVS.

De Indiarum jure, & Catholicorum Regum
obligationibus circa illarum administrationem.

*Cùm prædictum Ius amplißimæ hujus disputationis fundamentum
sit, ab illo exordium est optimo iure captandum: Vnde pro eo rogat*

CAPVT I.

An liceat de jure Indiarum in Catholicis Regibus dubitare.

§. I.

Negatiua Positio stabilitur.

STat pro illâ id, quod[14] Cardinalis Bellar[15]minus *Tractatu de Potestate Pontificis in temporalibus,* sapienter ostendit, præser[16]tim *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 etiam[17] Diana strenuè & eruditè decertat *Parte* 1. *Tract.* 2. *Resolut.* 122. *&* 123. *Parte* 10. *Tract.* 6. *Resolutione vnica & Tract.* 17. *Ac primò ex additis, Reso*[18]*lutionibus præsertim* 8. *&* 9. Vbi plurimos congerit huic dogmati suffragantes. Et vt alia præteream, veritatem istam luculenter probandam, & definiendam assumpsit Bonifacius Octauus in Extrauag. *Vnam Sanctam,* in quâ,[19] cùm de duobus gladiis, spirituali scilicet & materiali, quibus instructa Ecclesia, elegantem   discursum præmisisset, sic concludit: *Porrò subesse Romano Pontifici, omni humanæ creaturæ,*[20] *declaramus, dicimus, definimus, & pronuntiamus omnino esse de necessitate salutis.* Sic ille: vbi sensus est juxta præmissa; subesse inquam omnem humanam creaturam, non solùm spiritualiter, sed temporaliter: de spirituali enim[21] dubium non suberat, cùm Christus Diuo Petro omnes oues commiserit; pro quo *ibidem* Pontifex sic locutus: *Dicente Domino ipsi Petro: Pasce oues meas. Ioan.* 21. *v.* 17. *meas* inquit[22] *& generaliter, non singulariter, has vel illas: per quod commisisse sibi intelligitur vniuersas.* Ita ille: vnde quod specialiter suâ definitione constituit, adtemporalem pertinet potestatem: ex quo orta in Rege Franciæ, & regnicolis illa suspicio, quam Clemens Quintus suâ Constitu[23]tione submouit, declarans per definitionem & declarationem dictam, nullum Regi & regno præjudicium generari; neque ampliùs Pontifici subjectum, quàm ante prædictam declarationem & definitionem. Et meritò quidem: nam Bonifacius suâ declaratione & definitione, nouum jus non statuit,

23

24

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

Diego de Avendaño

præseferant, eorum etiam more vestiti, ad quos debet etiam cura se Protectoris extendere si ipsi velint; nam id fortè refugient, vt in suis contractibus agere liberiùs possint, & domicilium pro libitu commutare. Quod autem de Mulatis dictum, de liberis tantùm procedit: si enim mancipia sint, quia ex captiuâ nati, idem est quod de Æthiopibus asserendum; quando, etiamsi liberi essent, eorum deberet attendi paupertas; quæ quia communiter in prædictis occurrit, juxta eam judicandum; cùm aliàs paupertas ex abjecta origine incre-[41]mentum miseriæ dispiciatur; licet se solo non sufficiat: sunt namque multi, sicut & Mestizij adeò lautè viuentes, & in honore habiti, vt nullus eos miserabiles reputet, neque in eorum litigiis ad commiserationem peculiarem miserabilibus personis debitam, moueatur. Et quidem viduas, cùm diuites sunt, multi inter miserabiles personas numerandas negant, sicut & pupillos, vt videri potest apud Couarruuiam *Cap.* 6. citato. §. *Primus casus.* Quis enim miserabiles dicat diuites, id habentes, quo miserias possint omnes propulsare? Pecuniam inquam, de qua recens, sed elegans Poëta:

Me diua omnipotens, cusa, effigiata, rotunda.

Cujus & Hispana versio illa.

Aquella, si bien vndida Diosa, que todo lo puede,
Redonda, para que ruede, Porque se abata, batida.

Quòd si aliter dicendum, vt apud eumdem alij, de quo & *suprà num.* 154. ideò est quia in lege ipsâ expressum, in qua de viduis & pupillis mentio, vnde beneficium hoc non debuit eisdem, licet ratio pro ipsis non adeò esset vrgens, denegari. In casu autem nostro nulla est lex, quæ prædictos exprimat, & alioqui ratio non suffragatur ipsis: ex quo fit nullum futurum judicem, qui vti priuilegiatis fauendum in litigiis arbitretur.

[sect. 185] Quòd si fortè aliquis sit, per me certè licebit, & tueri se poterit ex eo quòd tales Pontificiis priuilegiis Indorum ac Æthiopum in ordine ad res spirituales, seu negotia animæ perfruantur, de quo dictum *Titulo 5. numer.* 134. quamquam id non adeò solidum sit: pro extensione enim præfatâ stat Pontificum expressa in quibusdam declaratio, vnde ad alia fit verosimilis ampliato. Quod secus in iis accidit, de quibus agimus: nullius enim legis, aut peculiaris rescripti pro eo indultum reperitur. Sed quia simile extat aliquid, vt in quibusdam aliis accidit, poterit tolerari, & pupillorum ac viduarum diuitum exemplum adscisci in fauorem prædictorum.

§. VII.

An ex præfatis origo puritati sanguinis obstet ad officia aliqua, & honorarios titulos requisita.

[sect. 186] DE hoc Dom. Solorzanus *citato Cap.* 27. *numer.* 77. *& seqq.* & in Politica *Cap.* 29. *pag.* 242. *Col.* 1. *& seqq.* 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, essetque 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 promiscuè loquitur; cùm 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

707

Thesaurus Indicus, Vol. 1

ex mancipiis descendunt, quorum ante captiuitatem genus probari nobile nullo modo potest, vnde & ceteris paribus Indigenis posthabentur.

[sect. 187] Quia verò in Societate nostra statutum de puritate dictâ extat *in Congregat.* 5. *Decreto* 52. *& 53. & in* 6. *Decreto* 28. obseruare juuat eos, qui ex Indorum genere nobiles vulgo habentur, nullum habere inde impedimentum. Sic *citato Decreto* 28. secus si ignobiles, vsque ad quintum gradum inclusiuè, ex eodem de Tarteris, & aliis ita concludente: intra quos tamen gradus dispensare R. P. Generalem posse, ibidem habetur. Quòd autem in eo Decreto Indica sit progenies inclusa, ex eo constat, quia minimè exclusa comperitur. Vbi enim? & quia quando opus est, ad eumdem 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 Æthiopibus certius multò.

[[42]] INDEX ADDITIONVM AD TOMVM PRIOREM THESAVRI INDICI. Numeri marginales hic tantum indicantur.

A

Æthiopes.

GAudent Indorum priuilegiis quoad temporalia, quia miserabiles personæ. 182. & 183
Ab ipsis origo an obstet sanguinis puritati ad multa requisitæ. 186

Alexander VI.

Illius defensio, & singularis deuotio erga Eucharistiam. 2

Angelus Maria Verricelli.

Illius erga Indiarum expugnationem & donationem Pontificiam asserta excutiuntur, cum donationis eiusdem defensione. 3. & seqq.

B

Bellum.

AN inferri posîit nolentibus Euangelij Prædicatores admittere. 8

Beneficium.

Circa Ecclesiastica, *vide* v. Electio.

708

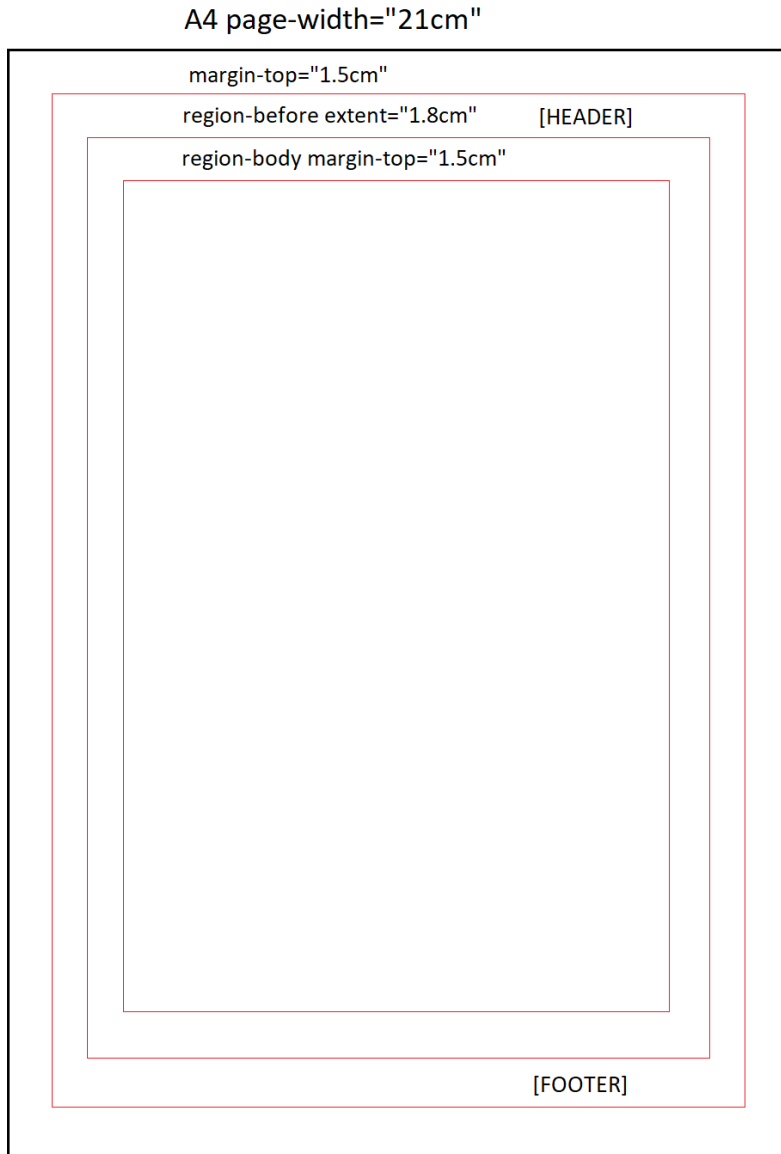● The publication concludes with endnotes.

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

We do not follow the canons of Western book page design, which places the center of the page area above the center of the page and where the gutter margin is traditionally narrower than the fore-edge margin[7]. This is because the publication is supposed to be printed or viewed as A4 pages (page-height="29.7cm" page-width="21cm") on a regular printer or monitor. The XSL-FO "page masters" prescribe three types of margins on each page side, two of which are not printable, and one containing a header (with the title and

---

[7] 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.

the author) and a footer (with page numbers) elements. Here an approximate layout of the page with the dimensions of the top margins:

A4 page-width="21cm"

margin-top="1.5cm"

region-before extent="1.8cm"   [HEADER]

region-body margin-top="1.5cm"

[FOOTER]

It is quite useful to define the print area as a one-cell table, the borders of which can be activated in debugging sessions to control the margins and other layout features.[8]

---

[8] A detailed description of the XSL-FO template is in preparation.

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.
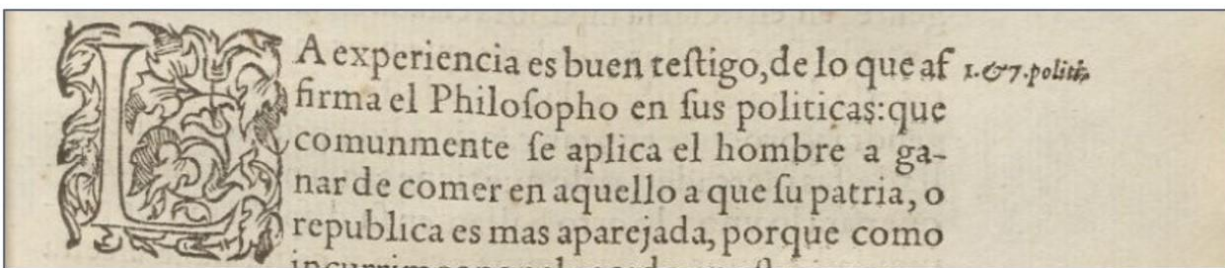
A. Order and position of XML elements

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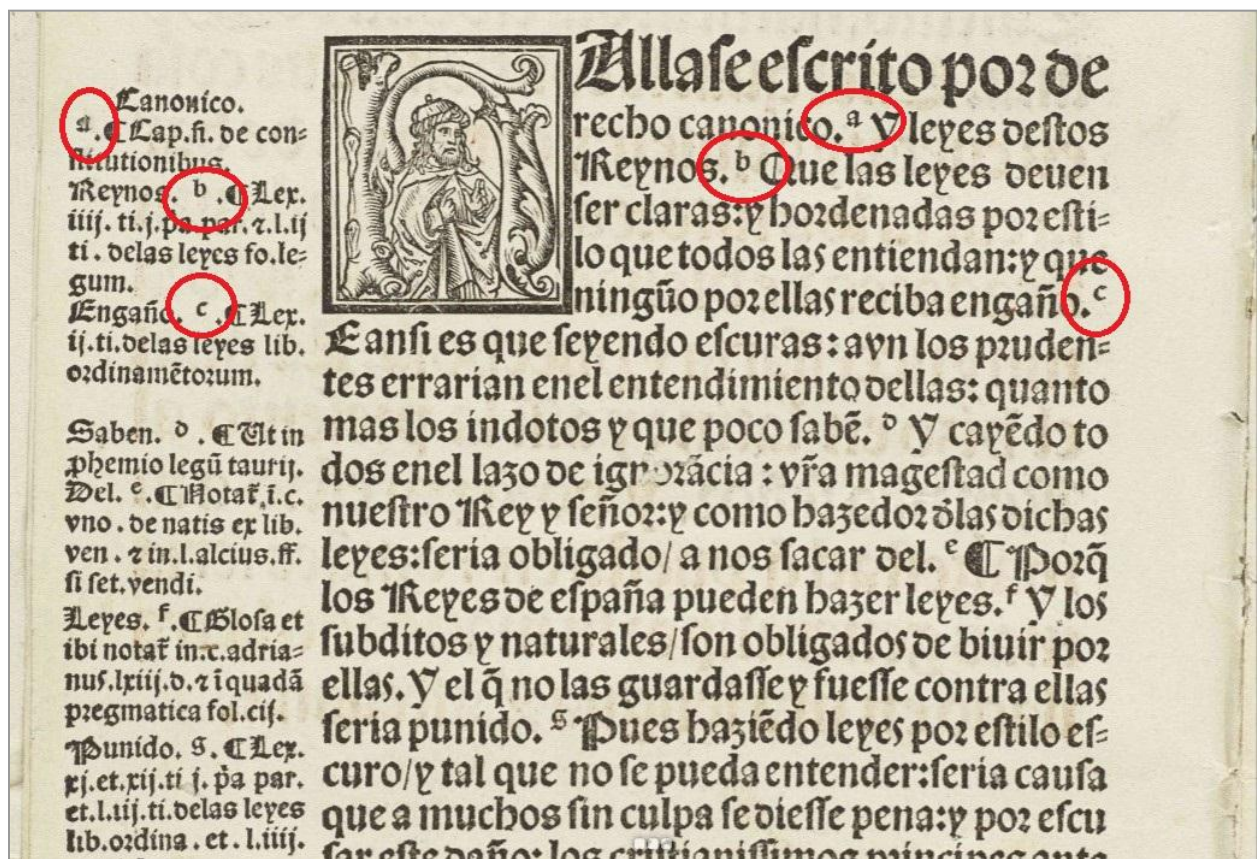
Sometimes the anchor is missing in the main text, and the note is just located next to the line it refers to. In XML this corresponds to the position before the <lb> element. The number of marginal notes in one document can be up to 8,000.

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. The point where they "float" is determined by the line beginning (<lb>) in XML, 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 normalize, i.e. unify the representation of anchored and unanchored notes – they all get a numerical mark and anchors in the main text in PDF. In addition, we place all the notes as endnotes at the end of the text, creating cross-references between them and anchors in text.

The side effect came to light with non-anchored notes situated at line endings with a word break. In this case the anchor appears in the middle of the word. From the PDF perspective it is desirable if marginal notes were encoded *before* the word beginning in XML[9].

---

[9] Considering the number of volumes already digitized and currently in production by the external transcription providers this change is now too labor-intensive to be implemented. This highlights again the necessity of visualization routines already at the stage of data model conceptualization.
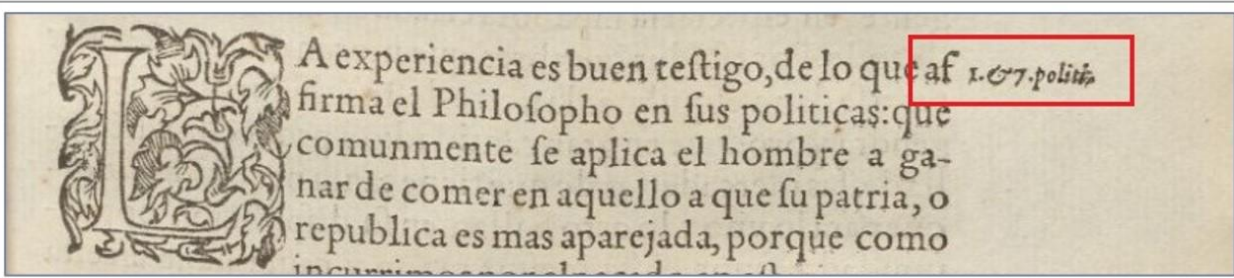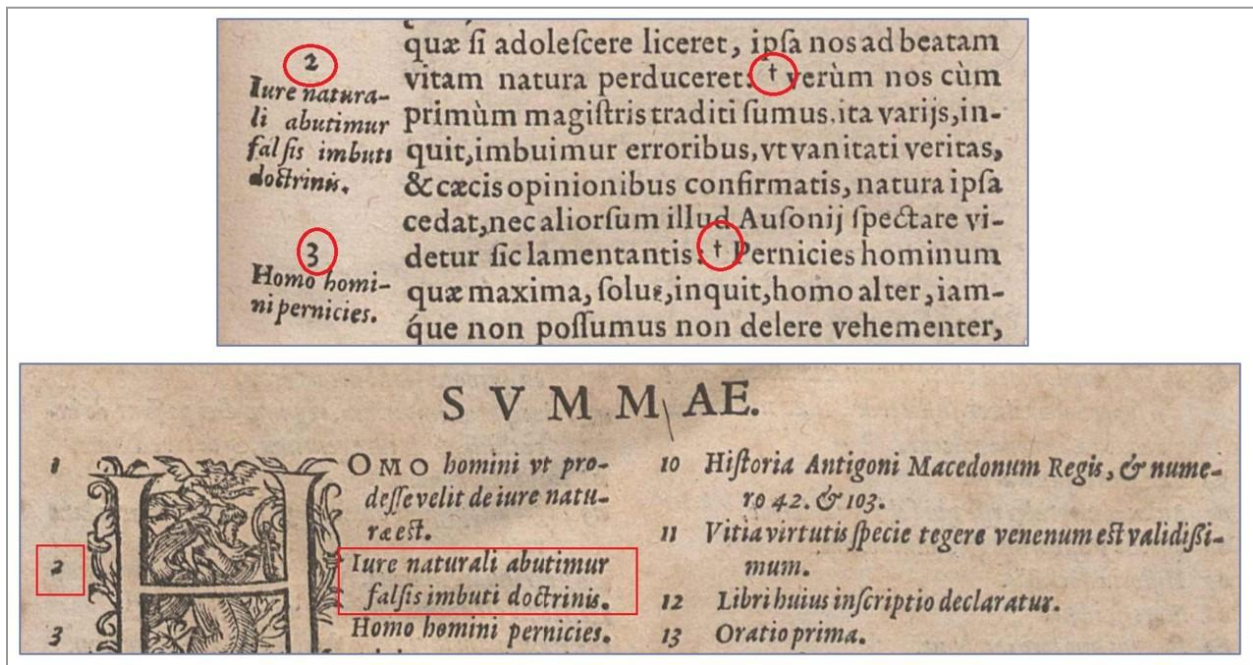
```
<hi rendition="#initCaps">L</hi>A experiencia es buen testigo, de lo que
<note place="margin" anchored="false" xml:lang="la" xml:id="xx">NOTE</note>
af<lb break="no" rendition="#noHyphen" xml:id="W0007-00-0035-lb-0006"/>firma
el Philosopho en sus politicas: que
```

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 essential component – an item in summary which they are linked to.

The problem with the word break also applies to milestones. Non-anchored milestones can split the word in the PDF edition if in XML they are encoded at word break. The same encoding recommendation applies to milestones as to the marginal notes – they should be placed before the word start and in clear cases at the beginning of the paragraph they refer to.

Title page and its constituents

The PDF printout can be used to test rules regarding the order of XML elements. For example, an item will not be rendered in PDF if it is located in a different position in the hierarchy than prescribed by the template. Alternatively, the template can be used to signal a changed structure. In this title page, for example, 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.

B. Cross-referencing

The PDF layout is a quite 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. This exceptional behavior was not caught up by Schematron, because it is a collateral bug of another improvement.

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 .

Another aspect is the actual design of a reference. Namely, which part of the string should serve as an anchor? "Summary" items can link just the page numbers (of the original), part of the title, or both. However, the original pages may be wrongly numbered and the print edition has its own page numbers. Therefore, referencing the whole item – as the element least susceptible to the output format – is preferable.


C. 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 **ſ** (long s). In addition, Latin, Greek, Hebrew, and Arabic characters occur and need special fonts to be rendered properly. PDF can thus be used as a handy copy-editing and proofreading tool.

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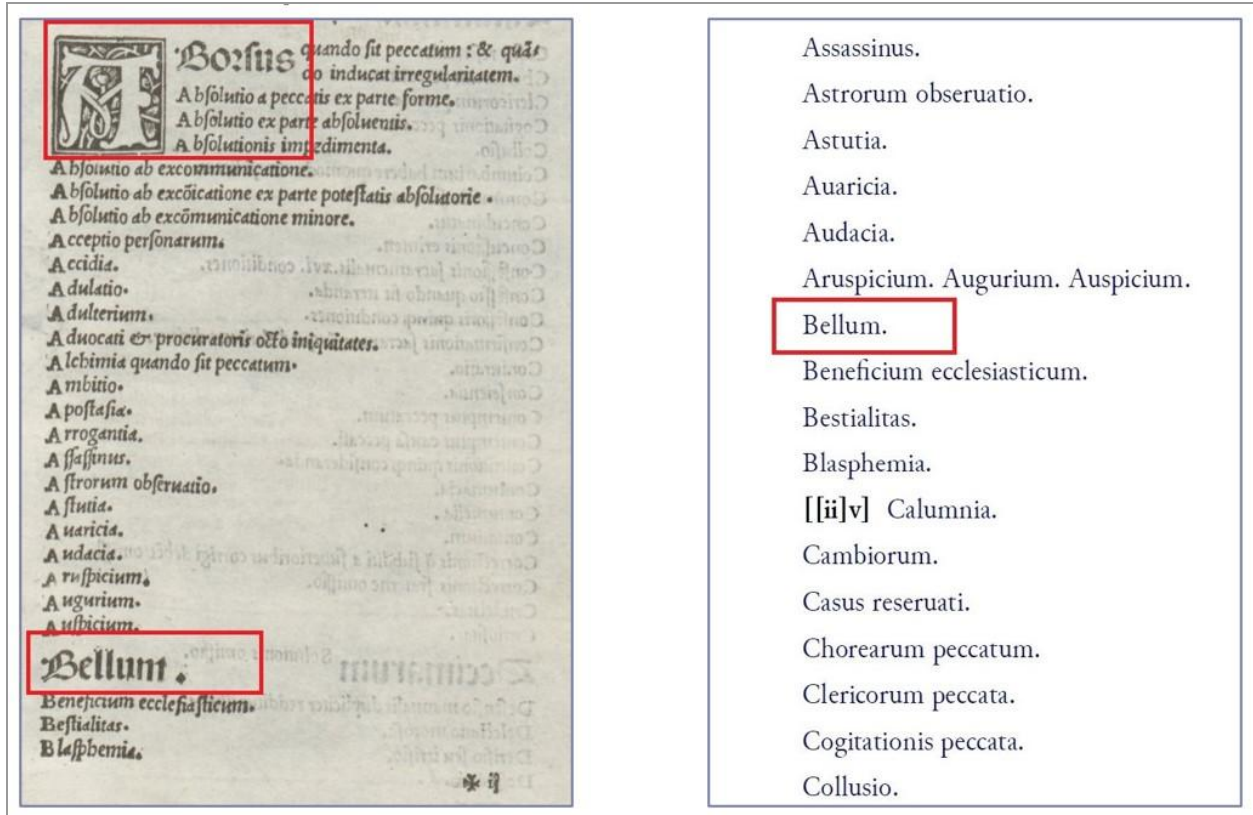
In the following example two functionally identical passages have different encodings, the second one marked as header .

[[v]] **FACVLTAS 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 Aquauiua 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 Vallisoliti vigesima quinta die mensis Augusti Anni 1608.

C. DE LOS COBOS.

[[vi]] **FACVLTAS 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

The most important concept visualized by PDF is the consistency of the encoding – the same principle should apply to the whole work.

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 issues raised depend on the definition of an "error", which in the context of scholarly digital editions is a rather challenging task. In the classical 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 classical 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[10]. 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 pipeline.
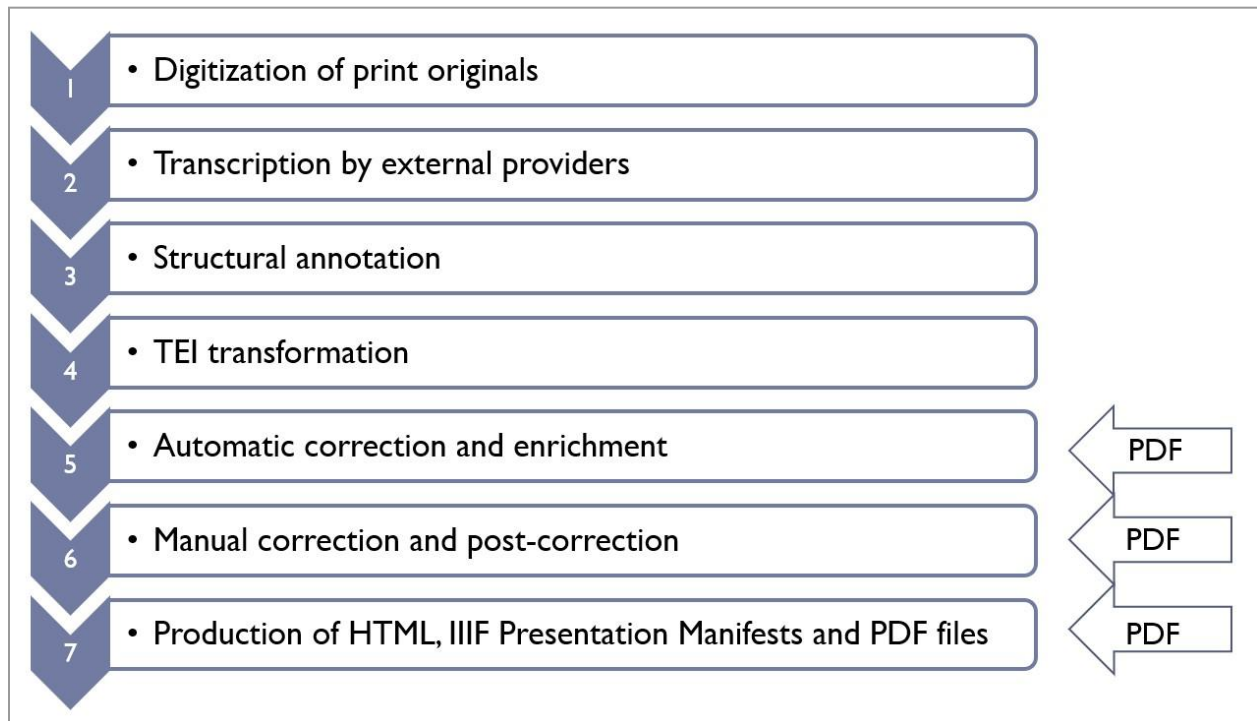
_____

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).
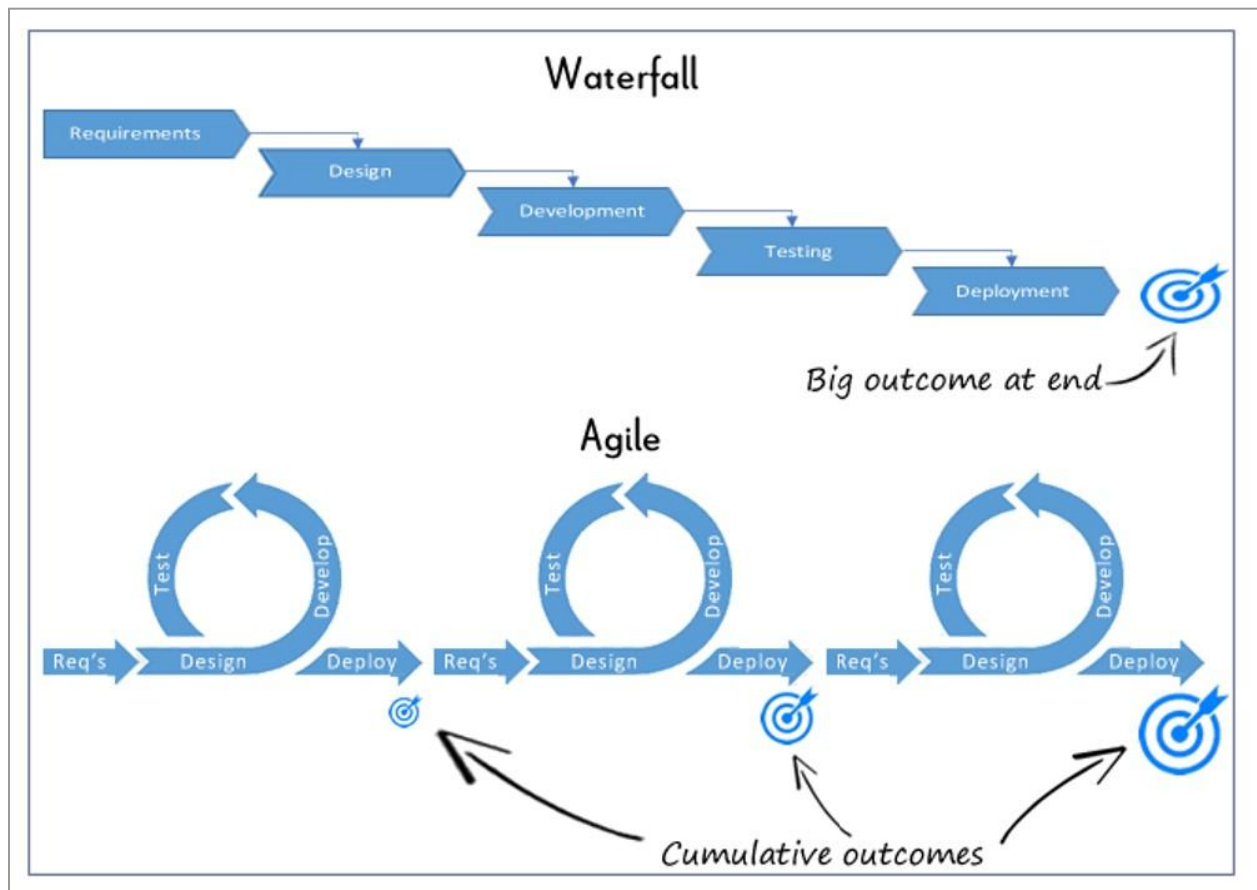

Image source: http://crmsearch.com/images/agileandwaterfall.gif

PDF creation for TEI editions 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. In this model researchers function as testers providing the 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 deep 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.

Maxim N. Kupreyev

The School of Salamanca Project

Bibliography

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.