

Data Entry Specs 1.0.1

Wolfgang Schmidle, Klaus Thoden, Malcolm D. Hyman
Max Planck Institute for the History of Science, Berlin, Germany
2nd October 2008

Contents

File Conventions	3
1 General Markup	3
1.1 Pages	3
1.1.1 Page breaks, Page Numbers and Running Heads	3
1.1.2 Catchwords and Signatures	4
1.2 Text Blocks	4
1.2.1 Headings	4
1.2.2 Paragraphs	5
1.2.3 Block Quotations	6
1.2.4 Footers	6
1.3 Columns and Tables	7
1.3.1 Columns	7
1.3.2 Tables	8
1.4 Notes	9
1.4.1 Marginal Notes	9
1.4.2 Footnotes	10
1.4.3 Anchored Marginal Notes	11
1.5 Printed Images	11
1.5.1 Figures	11
1.5.2 Drop Caps	12
1.6 Handwritten Notes	13
1.7 Unreadable Text	14
1.7.1 Characters You are Unsure About	14
1.7.2 Unknown Characters	14
2 Latin Alphabet	15
2.1 General	15
2.1.1 Punctuation	15
2.1.2 Hyphens	15
2.2 Characters	16
2.2.1 Characters to be Typed Directly	16
2.2.2 Other diacritics	17

2.3	Type Styles	17
2.3.1	Italics	17
2.3.2	Bold Face	17
2.3.3	Small Caps	18
2.3.4	Subscript and Superscript	18
2.4	Latin Ligatures	18
2.5	Example	19
3	Greek Alphabet	20
3.1	Greek Punctuation	20
3.2	Greek Ligatures	20
4	Mathematical Symbols	22
4.1	Fractions	22
A	List of All Tags	24
B	List of All Trial-and-error Issues	25
C	To Do	25
C.1	Questions	25
C.2	Changes From 1.0 to 1.0.1	26
C.3	To Do for Version 1.1	26
C.4	To Do for Version 1.2	29
C.5	To Do for Version 2.0	29
C.6	To Do for Later Versions	29
C.7	Points for the Post-processing Stage	29
D	Colophon	30
D.1	Fonts	30
D.2	Macros for Environments	30

File Conventions

Save the text in plain text format (.txt) with Unicode utf-8 encoding. If the text is saved in more than one file, enumerate the parts, for example Euclid_part_001.txt, Euclid_part_002.txt, and so on. Create a zip archive from all files.

We will also need the list of unknown characters (see section 1.7.2). If the list is handwritten, scan it in and save it as PDF file, if possible.

If file names can be only 8 characters long, use part_001.txt, part_002.txt, and so on. If the plain text format does not work properly, use the rich text format .rtf instead of .txt. Only if both .txt and .rtf do not work, use the Microsoft Word format .doc. Do not use any compression format other than zip. If you cannot use zip, leave the files uncompressed.

1 General Markup

1.1 Pages

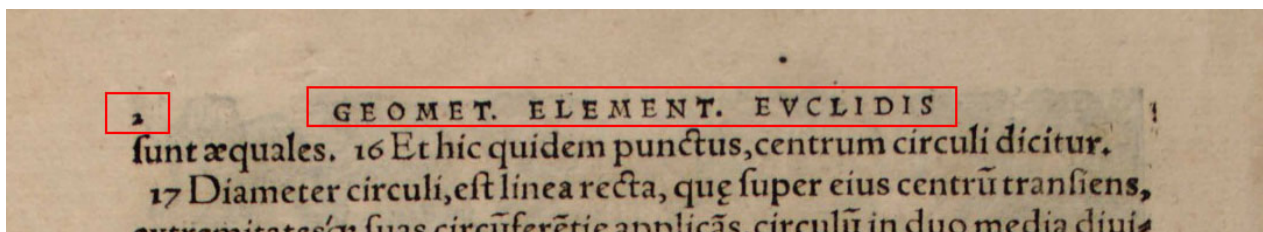
Type the entire content of one page, then go on to the next page. Do not mix the content of different pages.

1.1.1 Page breaks, Page Numbers and Running Heads

Page breaks are marked by <pb>. If the page has a page number, type it within the <pb> tag. Type the page number exactly as it appears in the book. If there is a running head on the page, it is marked by <rh> and </rh>. Type the running head immediately after the <pb> tag.

The position of the page number, e.g. at the top or bottom of the page, will not be encoded. Type the <pb> and <rh> tags before you type any content of the page. Do not type spaces within words. If there is a horizontal line below the running head, do not type it. A blank line may be inserted before the <pb> tag.

EXAMPLE 1



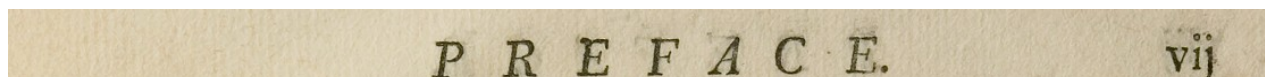
```
<pb 2><rh>GEOMET. ELEMENT. EVCLIDIS</rh>
```

```
$unt æquales. 16 Et hic quidem punctus, centrum circuli dicitur.</p>
```

(some untranscribed text)

→ For § and æ see section 2.2.1. </p> marks the end of a paragraph (section 1.2.2). The <p> for the beginning of the paragraph is on the previous page.

EXAMPLE 2



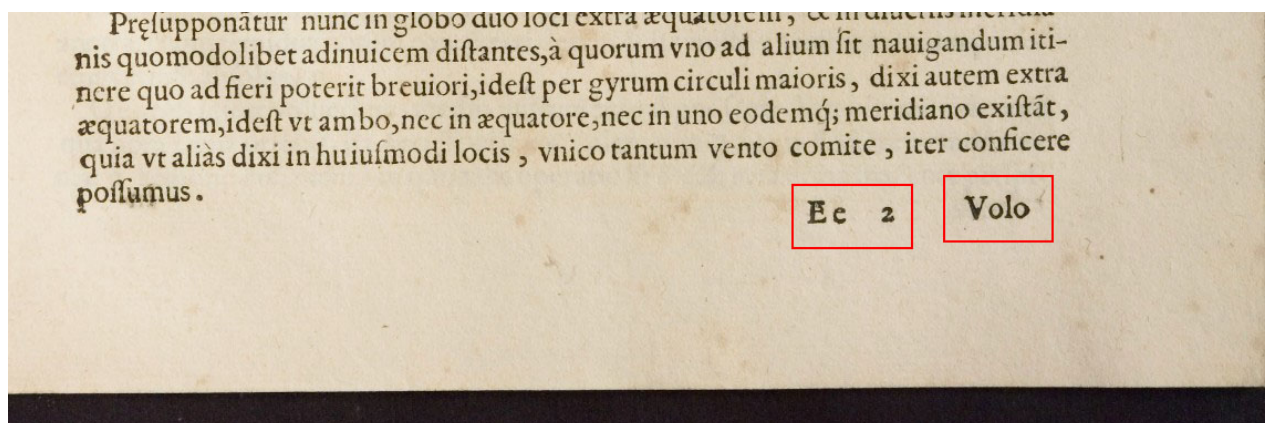
<pb vij><rh>PREFACE.</rh>

1.1.2 Catchwords and Signatures

Do not type catchwords and signatures.

In most cases, catchwords and signatures are at the bottom of the page.

EXAMPLE



(not transcribed)

The left rectangle contains the signature (“Ec 2”) and the right rectangle the catchword (“Volo”).

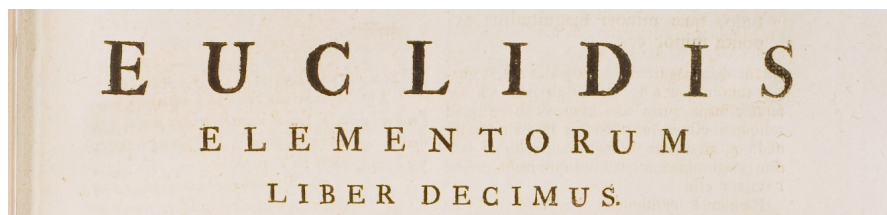
1.2 Text Blocks

1.2.1 Headings

Headings are marked by <h> and </h>.

All headings are tagged in the same way, regardless of the font size. Do not type spaces within words. If the text is centered, this will not be encoded.

EXAMPLE



<h>EUCLIDIS
ELEMENTORUM
LIBER DECIMUS.</h>

or alternatively, if you are unsure whether each line is a separate heading:

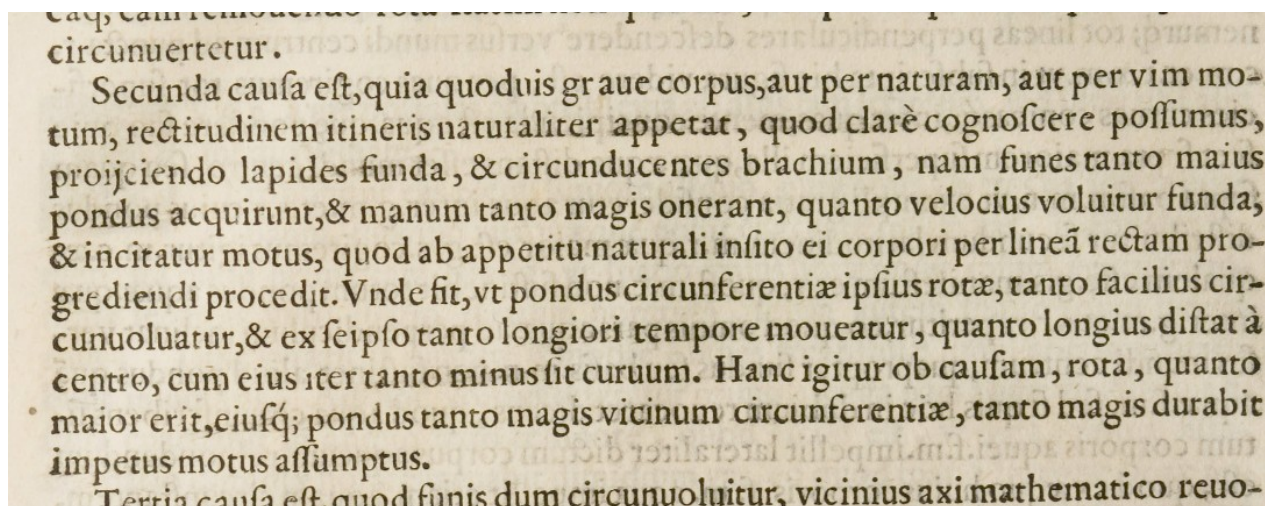
<h>EUCLIDIS</h>
<h>ELEMENTORUM</h>
<h>LIBER DECIMUS.</h>

1.2.2 Paragraphs

Paragraphs are marked by <p> and </p>.

Make sure that for each <p> there is a corresponding </p> somewhere. If the first line of the paragraph is indented, this will not be encoded. If the text is centered, this will not be encoded either.

EXAMPLE



(some untranscribed text)

<p>Secunda cauſa eſt, quia quoduis graue corpus, aut per naturam, aut per vim motum, rectitudinem itineris naturaliter appetat, quod clarè cognoscere poſſumus, proijciendo lapides funda, & circumducentes brachium, nam funes tanto maius pondus acquirunt, & manum tanto magis onerant, quanto velocius voluitur funda, & incitatur motus, quod ab appetitu naturali inſito ei corpori per lineã rectam progrediendi procedit. Vnde fit, vt pondus circumferentiã ipſius rotæ, tanto facilius circunuoluatur, & ex ſeipſo tanto longiori tempore moueatur, quanto longius diſtat à centro, cum eius iter tanto minus ſit curuum. Hanc igitur ob cauſam, rota, quanto maior erit, eiuſq; pondus tanto magis vicinum circumferentiã, tanto magis durabit impetus motus aſſumptus.</p>

(some untranscribed text)

→ For à and ā see section 2.2.1. For \`q see section 2.2.2. See also the example in section 2.5.

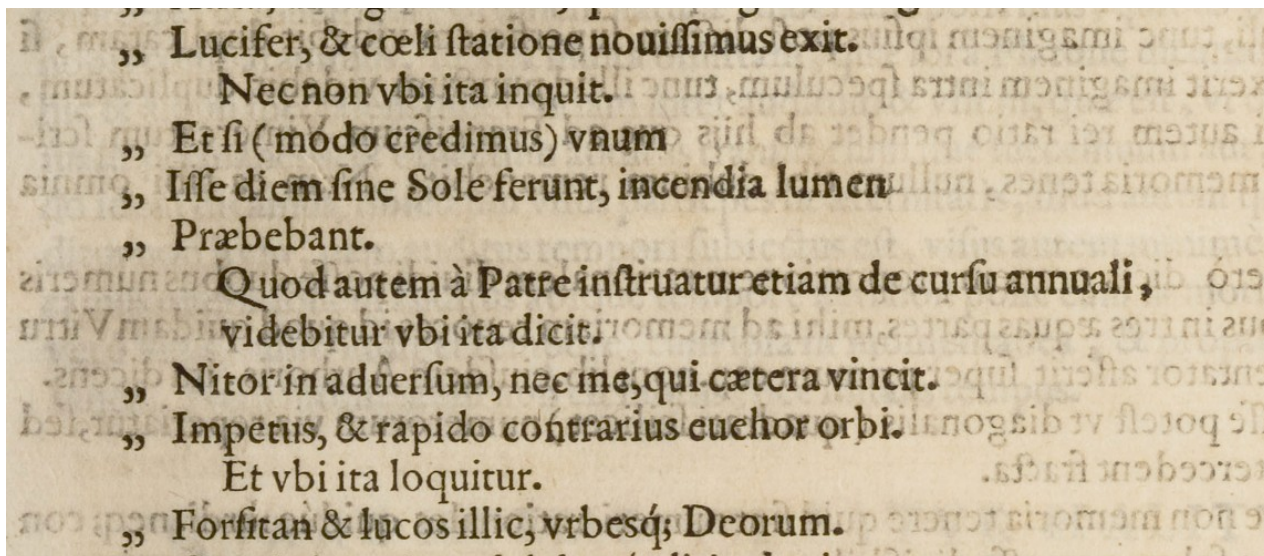
EXCEPTION: The <p> and </p> tags are not used in marginal notes (section 1.4.1) and footnotes (section 1.4.2). Headings (section 1.2.1) are marked by <h> </h> instead of <p> </p>. Block Quotations (section 1.2.3) are marked by <q> </q> instead of <p> </p>.

1.2.3 Block Quotations

A block quotation is marked by `<q>` and `</q>`. Do not type repeating quotation symbols.

The `<q>` and `</q>` replace the `<p>` and `</p>` tags.

EXAMPLE



```
<q>Lucifer, & cœli statione nouissimus exit.</q>  
<p>Nec non vbi ita inquit.</p>  
<q>Et si (modo credimus) vnum  
Iffe diem sine Sole ferunt, incendia lumen  
Præbebant.</q>  
<p>Quod autem à Patre instruantur etiam de cursu annuali,  
videbitur vbi dicit.</p>  
<q>Nitor in aduersum, nec me, qui cætera vincit.  
Impetus, & rapido contrarius euehor orbi.</q>  
<p>Et vbi ita loquitur.</p>  
<q>Forfitan & lucos illic, vrbes\'; Deorum.</q>
```

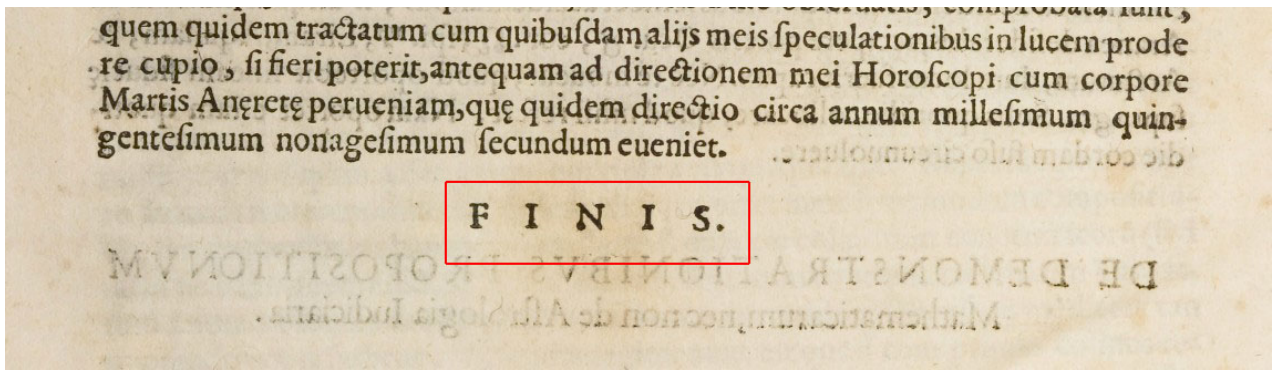
PLEASE NOTE: For inline quotations within a paragraph, type the quotation marks exactly as they appear in the text.

1.2.4 Footers

If you can identify a paragraph as a footer, use `<h>` and `</h>` instead of `<p>` and `</p>`.

→ `<h>` and `</h>` is the tag for headings (section 1.2.1).

EXAMPLE



`<p>(some text)`

quem quidem tractatum cum quibusdam alijs meis speculationibus in lucem prode re cupio, si fieri poterit, antequam ad directionem mei Horoscopi cum corpore Martis Aneret perueniam, quę quidem directio circa annum millesimum quingentesimum nonagesimum secundum eueniet.

`<h>FINIS.</h>`

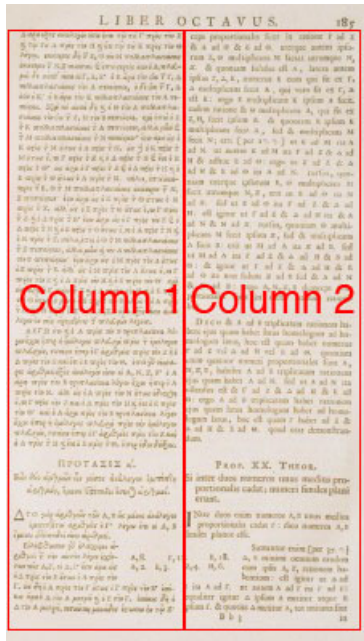
1.3 Columns and Tables

1.3.1 Columns

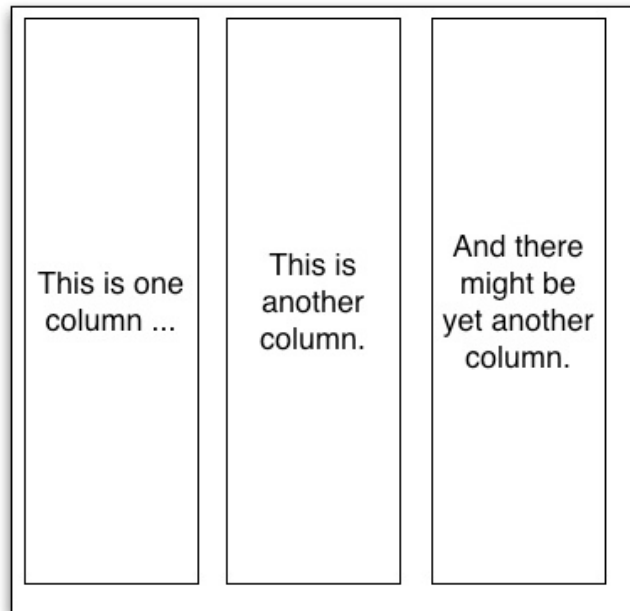
Columns are marked by `<col>` and `</col>`. The columns on a page are numbered from left to right.

EXAMPLE

(1) a real page:



(2) how to type columns:



(not transcribed)

PLEASE NOTE: The page number and the running head are not part of a column.

```
<col 1>This is one
column ...</col>
<col 2>This is
another
column.</col>
<col 3>And there
might be
yet another
column.</col>
```

PLEASE NOTE: If there is no running text in the columns, they may be not be separate columns, but a table (section 1.3.2). If in doubt, check the example there.

1.3.2 Tables

A table is marked by <tb> and </tb>. Use # as field separators. Do not type horizontal or vertical lines.

Do not treat the table columns as separate columns (section 1.3.1), i.e. do not type a whole column before you go on to the next column. If the whole table is in italics (section 2.3.1), indicate this in the <tb> tag, i.e. <tb it>.

EXAMPLE

Pag.	Lin.	Errata	Correcta	Pag.	Lin.	Errata	Correcta
3	29	equalis	aquali	158	26	versa	versam
8	35	maius	maior	158	26	fit	sint
9	15	in uitate superficialis, erit ac	inuitate superficialis erit, ac	162	22	cindenda	scindenda
11	1	proueniens	prouenientem	163	7	oppositus	oppositum


```

<tb it>Pag. # Lin. # Errata # Correcta
3 # 29 # æqualis # æquali
8 # 35 # maius # maior
9 # 15 # in vnitate $uperficialis, erit ac # in vnitate $upreficialis erit, ac
11 # 1 # proueuens # prouenientem
(some text)</tb>
<tb it>Pag. # Lin. # Errata # Correcta
158 # 26 # ver$a # ver$am
158 # 26 # $it # $int
162 # 22 # cindenda # $cindenda
163 # 7 # oppo$itus # oppo$itum
(some text)</tb>

```

If you are unsure whether some columns form a separate table or not, type it as one table:

```

<tb it>Pag. # Lin. # Errata # Correcta # Pag. # Lin. # Errata # Correcta
3 # 29 # æqualis # æquali # 158 # 26 # ver$a # ver$am
8 # 35 # maius # maior # 158 # 26 # $it # $int
(some untranscribed text)

```

PLEASE NOTE: If the columns contain running text, they may not be a table, but separate columns (section 1.3.1). If in doubt, check the example there.

1.4 Notes

PLEASE NOTE: Handwritten notes (section 1.6) are not transcribed.

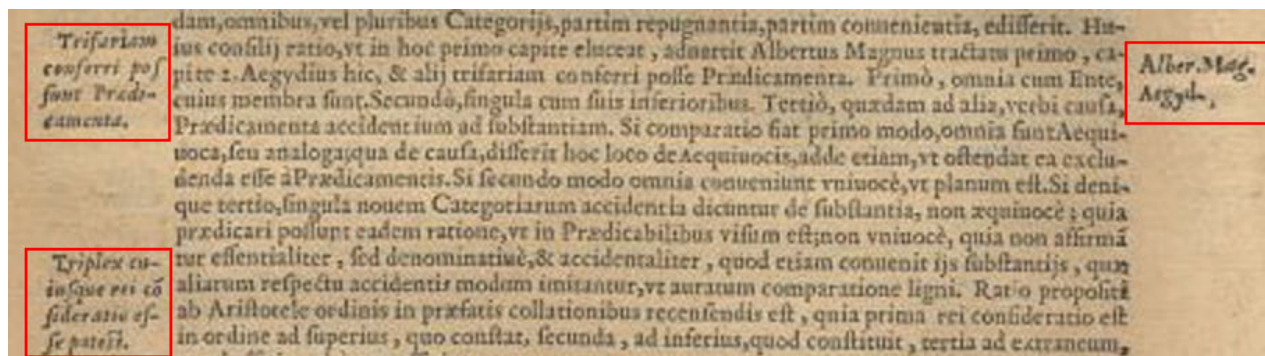
1.4.1 Marginal Notes

A marginal note on the left is marked by <mg1> </mg1>, and a marginal note on the right is marked by <mgr> </mgr>. It should be typed in separate lines, starting after the line it is the closest to.

Do not mark paragraphs within a marginal note, i.e. do not use the <p> tag.

→ For anchored marginal notes see section 1.4.3.

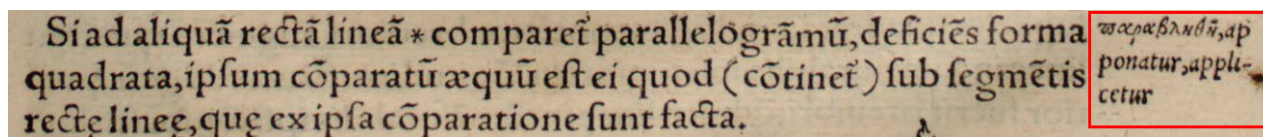
EXAMPLE 1



(not transcribed)

In example 1, there are two marginal notes on the left and one marginal note on the right.

EXAMPLE 2



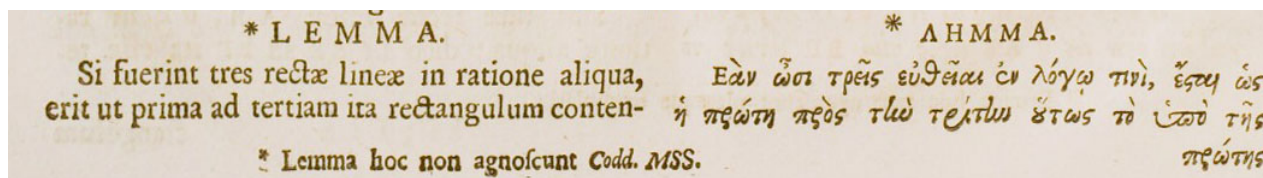
```
Si ad aliquã rectã lineã * compare\~t parallelogrãm\~u, defici\~es forma
<mgr>παραβληθῆ, ἀρ
ponatur, ἀρρί-
cetur</mgr>
quadrata, ipſum cõparat\~u æqu\~u eſt ei quod (cõtine\~t) ſub ſegm\~etis
rect\,e line\,e, qu\,e ex ipſa cõparatione ſunt facta.
```

1.4.2 Footnotes

Footnotes are marked by <fn> and </fn>. Type the footnote where it appears on the page. Insert the footnote symbol in the <fn> tag. The corresponding footnote symbol in the main text is marked by <n>. Insert the footnote symbol in the <n> tag.

Ignore the superscript of the footnote symbol. Use <n> in the main text only if you can identify the symbol as footnote symbol and if there is a corresponding <fn> </fn> tag somewhere. Do not mark paragraphs within a footnote, i.e. do not use the <p> </p> tag. Do not type horizontal lines.

EXAMPLE



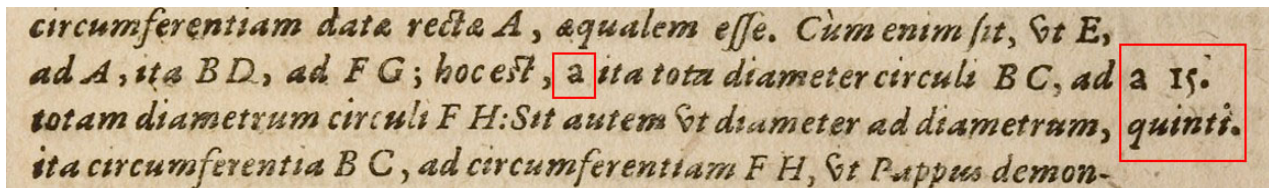
```
<col 1>
<h><n *> LEMMA</h>
<p>Si fuerint tres rectæ lineæ in ratione aliqua,
erit ut prima ad tertiam ita rectangulum conte-</col>
<col 2>
<h>ΔΗΜΜΑ.</h>
<p>Εάν ὡσι τρεῖς εὐθεῖαι ἐν λόγῳ τινι, ἔσται ὡς
ἡ πρώτη πρὸς τὴν τρίτην ἕτως τὸ ὑπὸ τῆς
</col>
<fn *>Lemma hoc non agnoscunt _Codd. MSS._</fn>
```

PLEASE NOTE: If the footnote continues on the next page, close the first part with </fn> before you go on to the next page (contrary to paragraphs). Mark the second part on the next page with new <fn> </fn> tags. If you remember the footnote symbol, type it again in the <fn> tag on the second page.

1.4.3 Anchored Marginal Notes

An anchored marginal note is marked by `<mg1>` `</mg1>` or `<mgr>` `</mgr>` according to its position, and typed in separate lines starting after the line it is the closest to (just like a normal marginal note). In addition, the anchor is treated like a footnote symbol, i.e. it is marked by `<n>` in the main text and it is written inside the `<mg1>` or `<mgr>` tag.

EXAMPLE



```
<p it>(some text)
circumferentiam data recta A, æqualem esse. Cùm enim sit, vt E,
ad A, ita B D, ad F G; hoc est, <n a> ita tota diameter circuli B C, ad
<mgr a>15.
quinti.</mgr>
totam diametrum circuli F H: Sit autem vt diameter ad diametrum,
ita circumferentia B C, ad circumferentiam F H, vt Pappus demon-
(some text)</p>
```

1.5 Printed Images

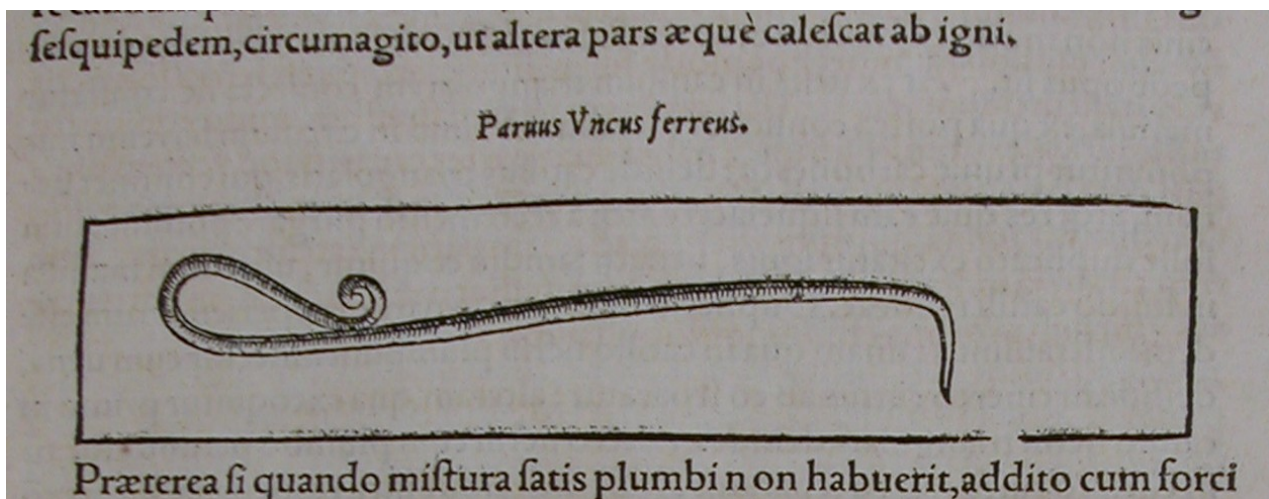
PLEASE NOTE: Treat handwritten figures in the same way as handwritten notes (section 1.6), i.e. simply mark them by `<hd>`.

1.5.1 Figures

Where a figure occurs in the text, put a `<fig>` tag in a separate line.

If you can identify a caption of the figure, mark it by `<cap>` `</cap>`. Type it in a separate line above or below the `<fig>` tag according to its position above or below the figure.

EXAMPLE I



<p>(some text)

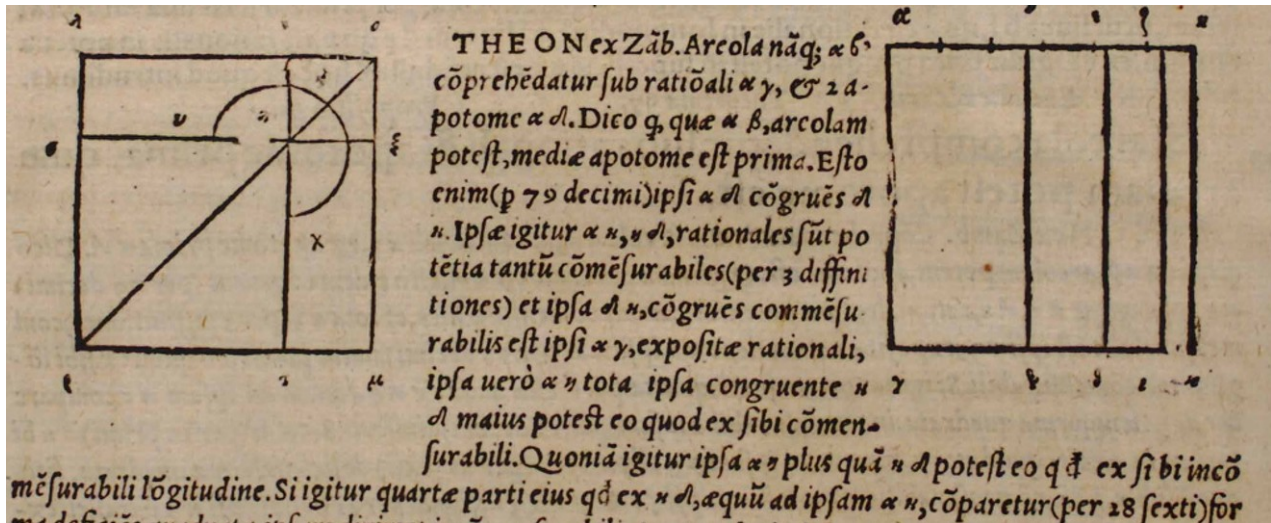
ſeſquipedem, circumagito, ut altera pars æquè caleſcat ab igni.</p>

<cap it>Paruus Vncus ferreus.</cap>

<fig>

<p>Præterea ſi quando miſtura ſatis plumbi non habuerit, addito cum forci
(some text)</p>

EXAMPLE 2



<fig>

<p>THEON ex Zãb. Areola nãq; α β,

<fig>

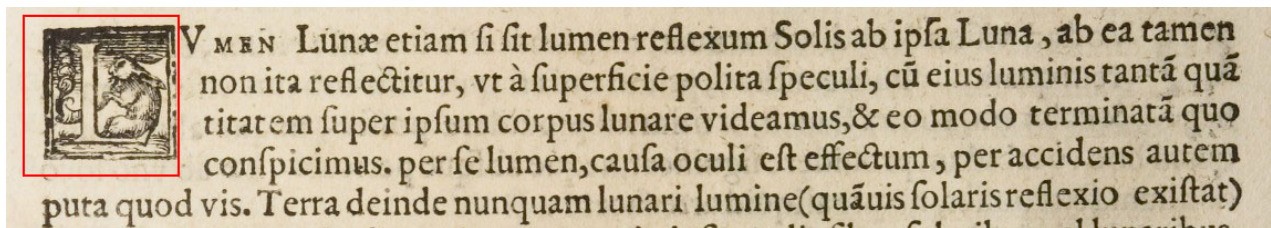
cōprehēdatur ſub ratiōali α γ, & 2 a-
potome α δ. Dico q, quæ α β, arcolam
poteſt, mediæ apotome eſt prima. Eſto
enim (p 79 decimi) ipſi α δ cōgru-
es δ
(some text)</p>

1.5.2 Drop Caps

Type a drop cap as an ordinary capital letter. Do not use the <fig> tag for ornamented drop caps.

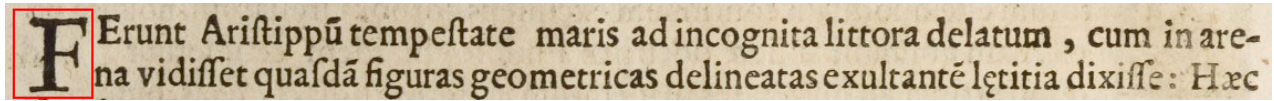
→ The <fig> tag marks figures (section 1.5.1).

EXAMPLE 1



`<p><sc>LVMEN</sc>` Lunæ etiam si sit lumen reflexum Solis ab ipsa Luna, ab ea tamen non ita reflectitur, ut à superficie polita speculi, cuius eius luminis tantâ quâ titatem super ipsum corpus lunare videamus, & eo modo terminatâ quo conspicimus. per se lumen, causa oculi est effectum, per accidens autem puta quod vis. Terra deinde nunquam lunari lumine (quâvis solaris reflexio existat) (some text)</p>

EXAMPLE 2



`<p>ERunt Arištipp\~u tempeštate maris ad incognita littora delatum, cum in arena vidisset quasdā figuras geometricas delineatas exultant\~e l\,etitia dixisse: Hæc` (some text)</p>

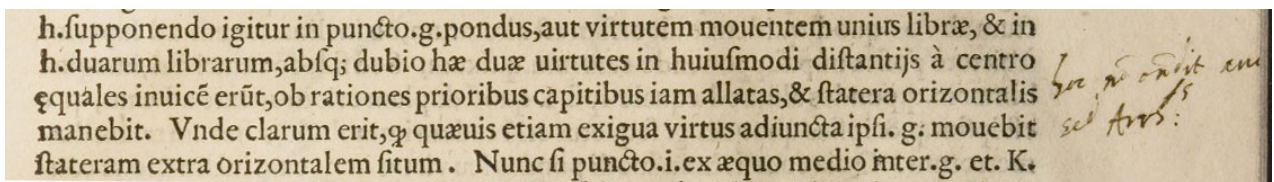
→ See also the example in section 2.5.

1.6 Handwritten Notes

Handwritten notes are marked by `<hd>`. Do not type the note itself.

The `<hd>` should be typed in a separate line, after the line of the main text it is the closest to.

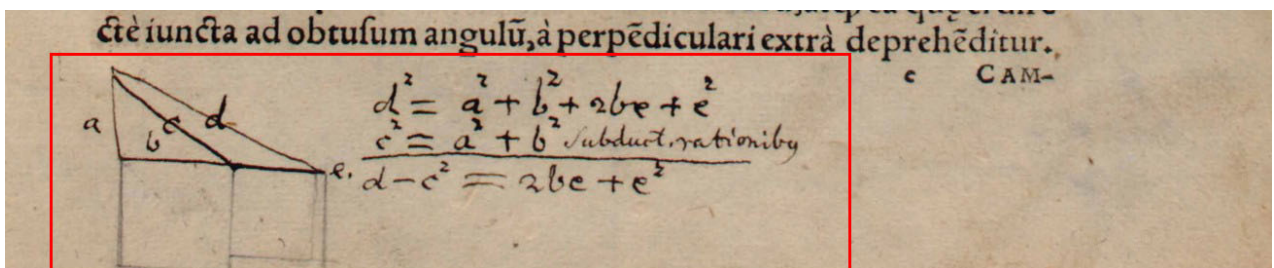
EXAMPLE 1



h. supponendo igitur in puncto. g. pondus, aut virtutem mouentem unius libræ, & in h. duarum librarum, absq; dubio hæ duæ uirtutes in huiusmodi distantijs à centro \,equales inuic\~e er\~ut, ob rationes prioribus capitibus iam allatas, & statera horizontalis

`<hd>` manebit. Vnde clarum erit, {quod?} quæuis etiam exigua virtus adiuncta ipsi. g. mouebit stateram extra horizontalem situm. Nunc si puncto. i. ex æquo medio inter. g. et. K.

EXAMPLE 2



`<p>(some text)`

ctè iuncta ad obtusum angul\~u, à perp\~ediculari extrà depreh\~editur.</p>

`<hd>`

→ The second example contains a signature and a catchword (section 1.1.2), which are ignored.

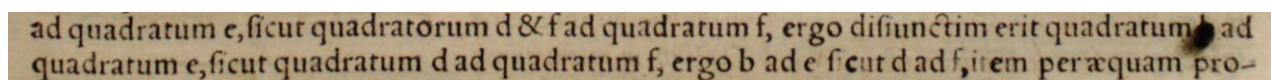
1.7 Unreadable Text

1.7.1 Characters You are Unsure About

If you are not sure about a character, type `<?>` after it. If you are unsure about a whole paragraph, type `<?>` directly after the `<p>` tag, i.e. `<p><?>`. Completely unreadable characters are typed as `@`. Use one `@` for each unreadable character, e.g. `unr@@dable`. If it is unclear how many characters are unreadable, use `<gap>` instead of `@`, e.g. `un<gap>le`.

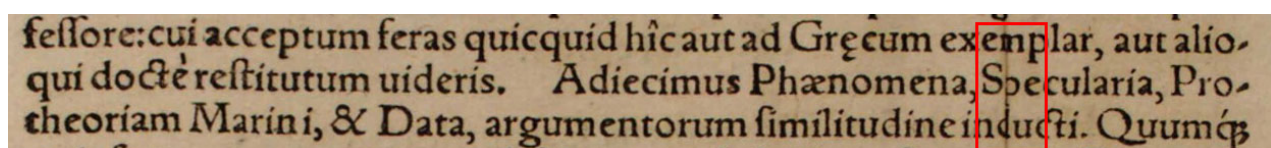
If you are unsure about a group of characters, for example a whole word, you don't need to type `<?>` repeatedly for every character.

EXAMPLE 1



ad quadratum e, sicut quadratorum d & f ad quadratum f, ergo disjunctim erit quadratum @ ad quadratum e, sicut quadratum d ad quadratum f, ergo b ad e sicut d ad f, item per æquam pro-

EXAMPLE 2



fe\$ore: cui acceptum feras quicquid hinc aut ad Gr\,ecum exem<?>plar, aut alio- qui doctè re\$titutum uideris. Adiecimus Phænomena, Sp@ecularia, Pro- theoriam Marini, & Data, argumentorum \$imilitudine ind<?>ucti. Quum{\`que}

PLEASE NOTE: For unknown characters please refer to section 1.7.2.

1.7.2 Unknown Characters

If there is an unknown character in the text, add it to the list of unknown characters. From then on, type its number whenever it occurs in the text.

Before you create a number for an unknown character, first check

- (1) whether it is in the list of ligatures (Latin: section 2.4, Greek: section 3.2),
- (2) whether it is in the list of characters to be typed directly section 2.2.1,
- (3) whether there is a standard for typing it (section 2.2.2), or
- (4) whether it is already on the list of unknown characters.

If the character is indeed unknown: Assign the number `<001>` to the first unknown character, `<002>` to the second unknown character, and so on. Do not assign the same number twice. Use this number to type the unknown character. Always use the same number if the same unknown character occurs again.

Make sure that for a given text there is a single list that contains all unknown characters, and that everyone uses this list. When the text is sent back to us, we will need a copy of this list.

→ For unreadable characters please refer to section 1.7.

2 Latin Alphabet

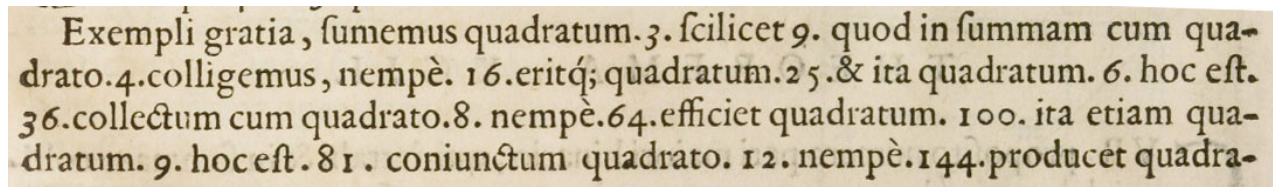
2.1 General

2.1.1 Punctuation

Type a space after the punctuation marks . , : ; ! ? Type a space before the opening brackets ([and after the closing brackets)]

Ignore superfluous and missing spaces before and after punctuation marks and brackets.

EXAMPLE



Exempli gratia, sumemus quadratum. 3. scilicet 9. quod in summam cum quadrato. 4. colligemus, nempè. 16. eritq; quadratum. 25. & ita quadratum. 6. hoc est. 36. collectum cum quadrato. 8. nempè. 64. efficiet quadratum. 100. ita etiam quadratum. 9. hoc est. 81. coniunctum quadrato. 12. nempè. 144. producet quadra-

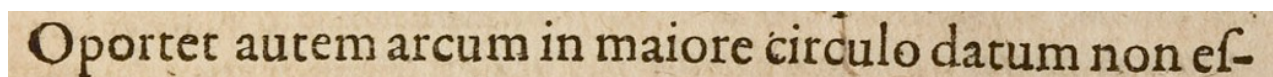
<p>Exempli gratia, sumemus quadratum. 3. scilicet 9. quod in summam cum quadrato. 4. colligemus, nempè. 16. eritq; quadratum. 25. & ita quadratum. 6. hoc est. 36. collectum cum quadrato. 8. nempè. 64. efficiet quadratum. 100. ita etiam quadratum. 9. hoc est. 81. coniunctum quadrato. 12. nempè. 144. producet quadra-
(some text)</p>

2.1.2 Hyphens

If the line ends with a hyphen, type it.

Hyphens at the end of a line can have different shapes. Always type a normal hyphen - regardless of its actual shape in the text. Do not insert a space at the end of the line.

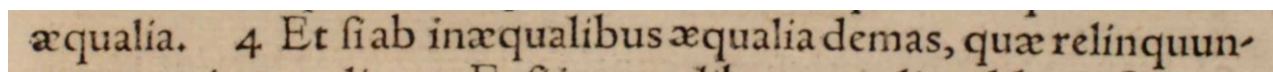
EXAMPLE 1



Oportet autem arcum in maiore circulo datum non ef-

Oportet autem arcum in maiore circulo datum non e\$-

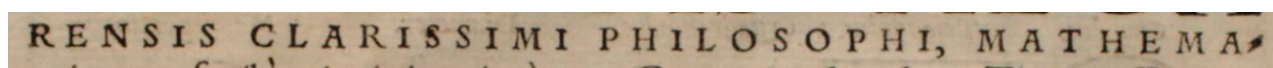
EXAMPLE 2



æqualia. 4 Et si ab inæqualibus æqualia demas, quæ relinquin-

æqualia. 4 Et si ab inæqualibus æqualia demas, quæ relinquin-

EXAMPLE 3



RENSIS CLARISSIMI PHILOSOPHI, MATHEMA-

RENSIS CLARISSIMI PHILOSOPHI, MATHEMA-

2.2 Characters

PLEASE NOTE: Section 1.7.2 explains how to deal with unknown characters. However, before you decide that a character is indeed unknown, first check (1) whether it is in the list of ligatures (Latin: section 2.4, Greek: section 3.2), (2) whether it is in the list in section 2.2.1, (3) whether there is a standard for typing it (section 2.2.2), or (4) whether it is already on the list of unknown characters.

2.2.1 Characters to be Typed Directly

The following characters with diacritics are to be typed directly:

Characters with acute accent

á (U+00E1) é (U+00E9) í (U+00ED) ó (U+00F3) ú (U+00FA)
Á (U+00C1) É (U+00C9) Í (U+00CD) Ó (U+00D3) Ú (U+00DA)

Characters with grave accent

à (U+00E0) è (U+00E8) ì (U+00EC) ò (U+00F2) ù (U+00F9)
À (U+00C0) È (U+00C8) Ì (U+00CC) Ò (U+00D2) Ù (U+00D9)

Characters with circumflex accent

â (U+00E2) ê (U+00EA) î (U+00EE) ô (U+00F4) û (U+00FB)
Â (U+00C2) Ê (U+00DA) Î (U+00CE) Ô (U+00D4) Û (U+00DB)

Characters with umlaut/diaeresis

ä (U+00E4) ë (U+00EB) ï (U+00EF) ö (U+00F6) ü (U+00FC) ÿ (U+00FF)
Ä (U+00C4) Ë (U+00CB) Ï (U+00CF) Ö (U+00D6) Ü (U+00DC) Ÿ (U+0178)

Characters with tilde

ã (U+00E3) õ (U+00F5) ñ (U+00F1)
Ã (U+00C3) Õ (U+00D5) Ñ (U+00D1)

Characters with cedilla

ç (U+00C7)
Ç (U+00E7)

Common ligatures

æ (U+00C6) œ (U+0153)
Æ (U+00E6) Œ (U+0152)

EXCEPTION: Type the character f (long s) as ſ.

2.2.2 Other diacritics

If a character is a combination of a known character and some known diacritic, type \, then the diacritic, then the character.

Use this method only if the character is not in the list in section 2.2.1.

EXAMPLE

q̇ q̈ q̃dratũ

\'q \`q \~qdrat\~u

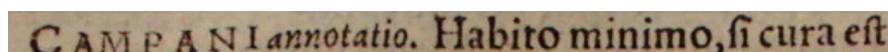
2.3 Type Styles

2.3.1 Italics

Use underscores _ _ to mark the beginning and end of words or whole lines in italics.

Encode only up to a few lines of text in this way. A whole paragraph (section 1.2.2) in italics is marked by *it* in the <p> tag, i.e. <p *it*>. If you mark a paragraph, do not mark the lines of the paragraph. The same applies to tables (section 1.3.2), i.e. <tb *it*>, and pages (section 1.1.1), i.e. <pb *it*>.

EXAMPLE



CAMPANI *annotatio. Habito minimo, si cura est*

<sc>CAMPANI</sc> *_annotatio_. Habito minimo, si cura est*

2.3.2 Bold Face

Use <bf> </bf> to mark bold face.

EXAMPLE

This is a **bold** statement.

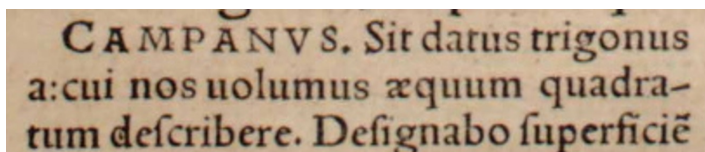
This is a <bf>bold</bf> statement.

2.3.3 Small Caps

If you can identify capital letters as small caps, use `<sc>` `</sc>`.

If the font size of the small caps differs from the rest of the paragraph, ignore this.

EXAMPLE



```
<p><sc>CAMPANVS</sc>. Sit datus trigonus  
a: cui nos uolumus æquum quadra-  
tum deſcribere. Deſignabo ſuperficiẽ  
(some text)</p>
```

2.3.4 Subscript and Superscript

Use `<_>` and `</_>` for subscript. Use `<^>` and `</^>` for superscript.

PLEASE NOTE: Numbers or other symbols in superscript may indicate a footnote (section 1.4.2). Check whether there is a corresponding footnote somewhere.

2.4 Latin Ligatures

Resolve common ligatures silently: *fi*, *fl*, *ffi*, *ffl*, *st*, *ct*. Resolve complex ligatures according to the provided list. Type `{` and `}` around the resolved characters of complex ligatures.

Do not resolve *æ* and *œ*, and type *f* as *ſ* (see section 2.2.1).

EXAMPLE 1

```
ff  fi  fl  ffi  ffl  st  ft  
ff  fi  fl  ffi  ffl  st  ſt
```

EXAMPLE 2

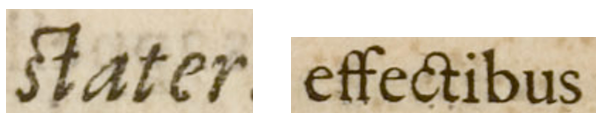


occaſione

eſſe

Sereniſſimo

poſteris



stater

effectibus

scientificè

scientificè

utriusq̃

q̃d

utrius{que}

{quod}

2.5 Example



<fig>
<h><fig>EVCLIDIS MEGA
RENSIS CLARISSIMI PHILOSOPHI, MATHEMA-
ticorum facilè principis, primùm ex Campano, deinde ex Theone Græco
cõmentatore, interprete Bartholomæo Zamberto Veneto,
Geometricorũ elementorum Liber primus.</h>
<h>Ex Campano, triplex principiorum genus.</h>
<h>Primùm, Diffinitiones.</h>
<p>PVnctus eſt, cuius pars non eſt. 2 Linea,
eſt Lõgitudò ſine latitudine: 3 cuius quidẽ
(some text)</p>

3 Greek Alphabet

3.1 Greek Punctuation

· (U+0387) Greek ano teleia

3.2 Greek Ligatures

Resolve ligatures according to the provided list. Type { and } around the resolved characters.

β β
ω π

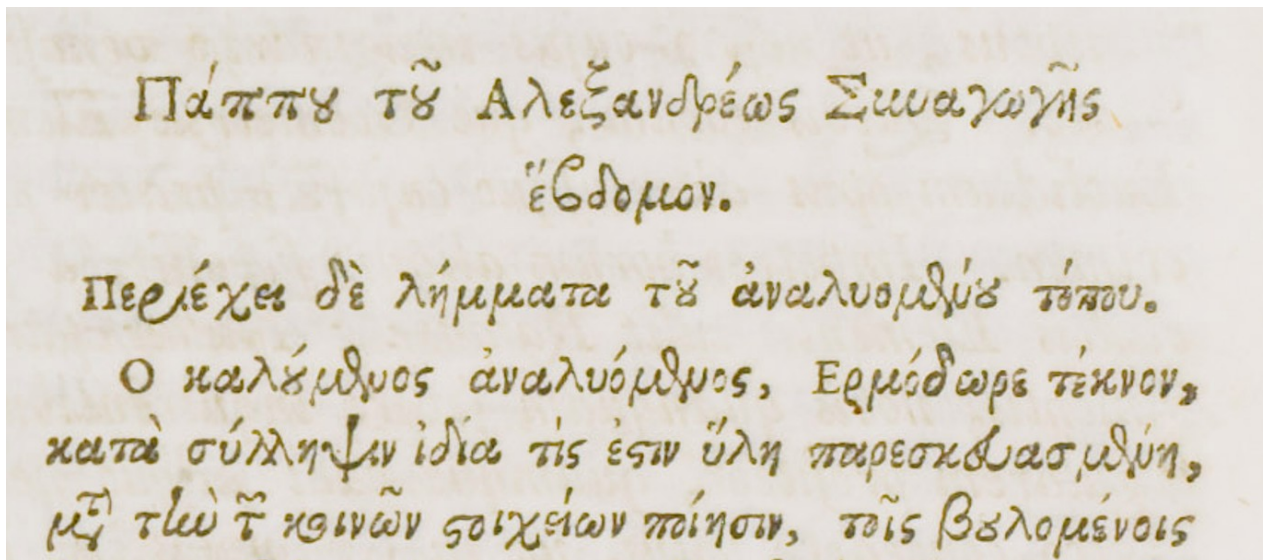
αι {αι} αῖ' {ᾰν} αῖ' {αύ}
ει {ει} εῖ {εῖ} εϋ {εν} εϋ {εϋ}
ηη {ηη}
λλ {λλ}
οξ {οξ} οξ̄ {ὀξ} ϣ {ου}
ρῖ {ρῖ}
σθ {σθ} σσ {σσ} στ {στ} σχ {σχ}
υυ {υυ}
χρ {χρ}

στί {στί} σο {στο}

γάρ {γάρ}
κατὰ {κατὰ}
μετὰ {μετὰ}
τῶν {τῶν}
καὶ {καὶ}
διὰ {διὰ}
ἐν {ἐν}

TODO: These are the ligatures from the philokalia package. The list may not be tailor-made for our needs (for instance, the τῶν ligature looks different), but it is a start.

EXAMPLE



<h>Πάππ{ου} τ{οῦ} Ἀλεξαν{δρ}έως Σ{υν}α{γω}{γῆ}ς
ἔβ{δο}{μο}ν.</h>

<p>Πε{ρι}έχ{ει} δὲ λήμμα{τα} τ{οῦ} ἀναλυο{μέν}{ου} {τό}{πο}υ.

Ο καλ{ού}{μεν}ος ἀναλυόμενος, Ερμόδωρε {τέ}κνον,
κατὰ σύλληψιν ἰδίᾳ τίς ἐστὶν ὕλη παρεσκευασμένη,
{μετὰ} τ{ῆν} {τῶν} {κοινῶν} {στοιχείων} {ποίησιν}, {τοῖς} βελομένοις

ἀναλαμβάνειν (έν) (γρ)αμμαίς δ{υν}αμιν εὔρε{τι}κ{ῆν}
{τῶν} {προ}τινο{μένων} {αὐ}{τοῖς} {προ}βλημάτων() (καὶ) εἰς
τ{οῦ}{το} μόνον (χρ)η{σί}μη κα(θε){στ}ῶ().

(some text)</p>

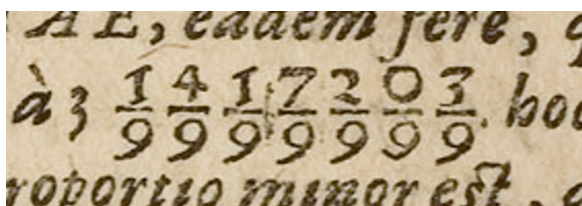
PLEASE NOTE: In Greek texts, the circumflex has two shapes (circumflex and tilde ~). Transcribe it as circumflex. The acute accent above ι may be vertical as in {τί}ς (line 5). However, type it as acute accent.

4 Mathematical Symbols

4.1 Fractions

Type fractions in one line. Use { / } to mark beginning, fraction line and ending.

EXAMPLE



à 3 {1417203/9999999}.

If you are unsure whether this is one fraction $\frac{1417203}{999999}$ or not, type it as separate fractions:

à 3 {1/9} {4/9} {1/9} {7/9} {2/9} {0/9} {3/9}.

A List of All Tags

section	tag	name	may contain
1.1.1	<pb>	page break	page number, it
1.1.1	<rh> </rh>	running head	it
1.2.1	<h> </h>	heading (footer)	it
1.2.2	<p> </p>	paragraph	it
1.2.3	<q> </q>	block quotation	it
1.3.1	<col> </col>	column	column number, it
1.3.2	<tb> </tb>	table	it
1.3.2	#	field separator	
1.4.1	<mgl> </mgl>	marginal note (left)	anchor symbol
1.4.1	<mgr> </mgr>	marginal note (right)	anchor symbol
1.4.2	<n>	footnote (main text)	footnote symbol
1.4.2	<fn> </fn>	footnote	footnote symbol
1.5.1	<fig>	figure	
1.5.1	<cap> </cap/>	figure caption	
1.6	<hd>	handwritten note	
1.7	@, <?>, <gap>	unreadable text	
<hr/>			
2.2.2	\' \' \~	character+diacritics	
1.7.2	<001>, etc.	unknown characters	
2.3.1	<u>word</u>	italics	
2.3.2	<bf> </bf>	bold face	
2.3.3	<sc> </sc>	small caps	
2.3.4	<_> </_>	subscript	
2.3.4	<^> </^>	superscript	
2.4	{quo}d, etc.	resolved ligature	
4.1	{ / }	fraction	

B List of All Trial-and-error Issues

- We will see what they will make out of marginal notes that consist of more than one paragraph.
- Can they identify footnote symbols and the anchors of anchored marginal notes?
- Does `_ _` for single non-italic words in paragraphs in italics work? (see To Do for version 1.1)
- if we introduce `</fig>`, will they identify figure captions?
- A page with two columns, and when one column is finished, the other columns uses the space of the other column as well.
- Paragraphs that begin on one page and end on the next page

Ask the company which characters they can type directly (probably at least the characters from the list in section 2.2.1). For example `ẽ` ?

C To Do

Leave the images in colour, but it should also work in b&w.

C.1 Questions

Section File Conventions: Does that make sense, and is the content appropriate?

Structure

In the new ordering, where does Fraktur fit in?

Does it really make sense to type drop caps as normal characters? Sometimes they are used as an additional structuring element below the section level (only in modern texts?)

`<n>` tag for footnote symbols in the main text: insert space before and after if there are spaces in the text (no rule, but how do we do it in our example)? If we use the modern way, e.g. comma, no space, symbol, space, we contradict our punctuation rule.

List of footnote symbols?

Figures, first example: Add a few lines of text at the top? That would make the placement of the `<fig>` tag more clear.

figure between two columns: treat as to the right of the left column? yes: do it (in an example?), without explaining it

Especially since we do not use Beta Code, we don't let them type e.g. `<lang greek>`, but include it later. Is that correct?

Latin Alphabet

Exception to punctuation rule? If you can identify an abbreviation without spaces after the periods, do not insert spaces. Example: i.q.d.et.i.d.q.

Replace the punctuation example (a weird table of Pythagorean triples)?

bold face for tags: I stick to the rule “everything that is not on the page will be printed in bold face”.

Table of characters to be typed directly: add more characters? (Probably we will stick to this core list.) Place the table more prominently? Larger font? Bold face?

`<p><it>` instead of `<p it>`? Problem with `<bp vij it>`.

Unicode Are characters like ã and ù included as single characters in Unicode? (Apparently yes.) If not, how does this “ã = q~” work?

q+circle is not supposed to be on the list of ligatures. How do I get q+circle in TeX? And how would they type it in China? The circle does not seem to be on the American keyboard.

Are there examples of Latin characters with two diacritics in the texts? (In general, there are examples for this, eg. “A with ring and acute”.)

C.2 Changes From 1.0 to 1.0.1

Added section “File Conventions”

New, more modular ordering of sections: General (was: Structural Markup), Latin Alphabet (was: Positional Markup), section Unknown Characters moved from Characters to Unreadable Text, Greek Ligatures moved into separate section Greek Alphabet, Mathematical Symbols as separate section.

Updated list of all tags

C.3 To Do for Version 1.1

include new notes

proofread all examples (again)

example: whole page of Latin text (i.e. finish typing the example at the end of “Latin Alphabet”)

(other diacritics: add an example with two diacritics?)

TODO: Problem of single non-italic words within a paragraph in italics: “Mark single non-italic words in a paragraph in italics by _ _” ? Is this binary switch too complicated? It was suggested that we try out _ _ and see what happens. (Alternative: | |.) (Leave out completely? If we leave it in, we need an example.)

TODO: Problem of single italic letters within a paragraph in italics: “Ignore the italics of single characters within a word in normal font”: leave out? (Does not apply to mathematical texts anyway?? The standard of using italics is not so old, I guess?)

Latin ligatures

TODO: list of Latin ligatures (from Paul Trzeciok?): que, quo

TODO: read archimedes.pen, it's on the wiki. Extract a sensible list of strange latin characters (many of them can be typed directly, e.g. radix). (IN: therefore, prime, double prime, pilcrow, plus, minus, dagger. OUT: white rectangle, horizontal ellipsis, spacing diacritics.)

TODO: If it can be typed as it is (character + diacritics), it should not be on the list (this belongs somewhere else!)

TODO: replace each image by a word in a Mac font that uses the respective ligature: DejaVu automatically does fi, fl, ffi, ffl. It also has the explicit ligatures st (latin small ligature st, U+FB06), long-s t (latin small ligature long s t, U+FB05). No ct? Is it possible to convince DejaVu to produce one? (And will there be any font that has the long-s ligatures fi ff ffi ft? Are DejaVu's fi ff ffi ft close enough? Fake them by fiddling with the space between the letters?) → NO faked ligatures!

TODO: List of complex Latin ligatures: Ligature, meaning (leave out?), encode as... Encode as <01> (or <001> ?) etc. according to this list

Greek ligatures

list Greek ligatures: 1 simple, 2 frequent, 3 rare (but see below)

TODO: aus der griechischen Seite ausschneiden??

TODO: do we need to provide a list of all (normal) Greek characters and accents/spiritus? two different sigmas. NO

TODO include allographs such as alternative beta (β, U+03D0), pi that looks like omega with a tilde, in this list of Greek characters? One could also include them in the list of ligatures, even though they are not ligatures. Start the list with “single character ligatures” ? YES, examples. Skip hard ligatures, they will be treated as unknown characters.

TODO: List of ligatures: finish on time, it does not have to be perfect at once. Start with a subset of Ingram's list (read his introduction!), use Faulmann and Wallace only if necessary. (→ check the whole Greek Euclid page: are all non-trivial ligatures contained in Ingram's list?) Use the (non-unicode) renaissance ligature font to avoid a handwritten list?

TODO list: high frequency ligatures: ου, μεν, stigma (or accept that they type end-sigma instead?), etc.

A table of two-letter ligatures. There are no images of simple two-letter ligatures. Images of difficult ligatures (in this table in bold face) can be seen below. TODO: sind die einfachen Ligaturen wirklich einfach??

	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο	π	ρ	σ	τ	υ	φ	χ	ψ	ω
α	αι	αυ
β
γ	γη	γρ	γω
δ	δι	δο	..	δρ
ε	ει	εν	ερ	..	ευ
ζ
η	ην
θ	θε
ι
κ	κο
λ	λλ
μ	μο
ν
ξ
ο	ου
π	πα	πο
ρ	ρι
σ	σι	σκ
τ	τα	τε	τι	το
υ	υν
φ
χ	χρ
ψ	ψι
ω

TODO: offenbar ist bei zwei-Buchstaben-Ligaturen der zweite Buchstabe ein Vokal (α, ε, η, ι, ο, υ, ω) oder ein Konsonant, der am Ende des Wortes auftreten kann (ν, ρ; Schluss-sigma ς wird nicht ligiert). Wenige Ausnahmen: λλ, σκ.

TODO: Problem z.B. υν: in der Ligatur ist der Akzent verschoben. Wird die Ligatur dadurch schwierig? Oder sagen: Akzent kann verschoben sein, Akzente nur auf Vokalen, d.h. ... ?

TODO: list: 1 (allographs) alternative beta, 2 (difficult two-letter ligatures) ευ, etc., 3 (ligatures of three or more letters) stigma-iota στι, stigma-omikron στο (aber wenn das die einzigen Beispiele sind, schummeln und zu 2 tun) 4 Abkürzungsligaturen (abbreviating ligatures): μετα, των, και, προ, γαρ, etc. (wo sortiert man Wort-Ligaturen von zwei-letter-Wörtern ein?)

1. alternative beta, 2. frequent proper ligatures, emphasis on hard and frequent

TODO: punctuation marks

TODO keep track of the origins of the examples (which book, which page), e.g. here: Euclid lat/gr, p.16. Make comment lines.

"Fractions are a kind of ligature"

TODO: In the last word, μέν is *not* ligated (the Chinese won't care) and the ου ligature is printed badly (if it is not a variant). Should we provide a list of common pitfalls? NO (maybe later)

C.4 To Do for Version 1.2

Fraktur

C.5 To Do for Version 2.0

Chinese

C.6 To Do for Later Versions

indexes: find examples

lists: find examples

section “Book-specific Specifications”, subsections “A Scheme for Book-specific Specifications” and “An Example”

Do we need to provide a list with `<01>` replacements for unusual footnote symbols?

Headings: maybe different rules for specific books

Tables: more precise rules for books with many (e.g. astronomical) tables

At the end: Make sure that example images and the corresponding text are on the same page. Introduce shorter tag names if there are no naming collisions.

modular structure: is it enough to say: For this specific book, ignore sections ... ?

add example for subscript and superscript

`<+>` and `<->` may be used for indicating different font sizes. Alternative: `<s+>`, `<s->` for superscript and subscript, `<fs+>`, `<fs->` for different font sizes.

TODO: Rules that also work for more complex formulas (i.e. as complex as the early typographers could manage; for example, no rule is needed for $\infty^{\frac{1}{2}}$, since in all likelihood it would have been set as $\infty_{\frac{1}{2}}$ anyway)

TODO (possible book-specific rule: space after sentence number)

C.7 Points for the Post-processing Stage

PLEASE NOTE: SR suggests some minimal post-processing when texts are added to the repository, namely a concordance of page image numbers and the page numbers as they appear on the page, so that the pages can be cited properly; and also checking for doubly used and missing page numbers. SR also mentioned old ECHO wish lists for the xml; where are these lists?

`<lang gr>`, etc., for example with Latin and Greek columns? If the standard language of the book is Latin, mark Greek columns only? Or is it easily insertable during post-processing? Is it needed at all, since they use unicode characters?

During post-processing, will we correct the main text according to errata lists? Or at least links? What about new typos in the errata lists? (There is one in the example!)

Is the open rule for figures, combined with some examples, sufficient? The main criticism of the archimedes style `<p>text<figure/></p>`, i.e. the figure tag within the paragraph, concerns only the post-processing stage.

When the last paragraph on a page continues on the next page, but there is a footnote at the bottom: In the post-processing stage, all notes are moved to the end of the text (with links).

The main reason for typing footnotes where they are (rather than marginal notes and LaTeX style footnotes) is that a footnote can continue on the next page. How do we encode the second part of the footnote? Ignore the problem in the specs, hoping that they will at least identify it as a (separate) paragraph, and merge the footnote parts in the post-processing stage?

A, B, Γ, Δ: A and B will inevitably be typed as Latin characters even though semantically they are Greek characters (same shape, but different Unicode codes). Are we fine with that? Problem for searching?

D Colophon

D.1 Fonts

Headings in Helvetica, normal text in DejaVu Serif (was Times), except small caps which are in Hoefler Text, Latin and Greek characters in normal text in DejaVu Serif, Latin and Greek examples in Courier New (was: DejaVu Sans Mono scale 0.8 / 0.9)

Problems of the DejaVu font:

- section 1.2.1: In the monospaced font for Latin, something like `DECIMUS.</h>` looks fine on screen, but on paper it looks as if there was a space between the period and the tag because the period is closer to the S.
- In the monospaced font for Greek, the epsilon and epsilon+accent look fine on screen, but are not printed (in the case of epsilon+accent, only the accent is printed). Only the small epsilon is affected!

Until these problems are sorted out, I will use Courier New as monospaced font again.

D.2 Macros for Environments

mainrule, clarification, note, exception, crossref

example, sampleImage, smallSampleImage

typeLatin, typeGreek, typeMath