

Instructions for transcription and XML encoding of the Herle letters

(Alison Wiggins and Robyn Adams, February 2005)

1. File types and file names

Each of Herle's c. 300 letters will be transcribed in text-only form then marked up in XML, according to the project DTD (herle.dtd, copy below). There will be one XML file for each letter. The name of each XML file will be the same as the id for each letter, as recorded in the Access database catalogue of letters. So, the letter with the id '001' will be stored in the file 001.xml, the letter with the id '002' will be 002.xml, and so on.

2. Declarations, root element and level-1 parent elements

Every letter transcript will begin with the same XML and DOCTYPE declaration, root element and level1 parent element:

[declarations:] `<?xml version="1.0" encoding="UTF-8" standalone="no" ?><!DOCTYPE herle SYSTEM "herle.dtd">`

[root element:] `<herle>`

[level-1 parent element:] `<herle_letter id="1">`

The 'id' attribute is mandatory. It's value will also be the same as the id number for the letter in question, as recorded in the catalogue of letters.

3. Level-2 parent elements

The transcription and editorial commentary for each Herle letter will be contained within three level-2 parent elements: `<editorial>`, `<address_leaf>` and `<letter_text>`. They must always appear in this order. The elements `<editorial>` and `<letter_text>` must occur once and once only each. The element `<address_leaf>` may occur once or not at all. No text or other elements can appear outside of these three level-2 parent elements.

4. Editorial Commentary `<editorial>`

4.1. The editorial commentary is included the head of the letter and defined by the `<editorial>` element. The `<editorial>` element has no child elements. The commentary is always written in coherent prose (not in note form). The following information will be included always in this order and always conforming with the format of the examples given below. Items 4.2 and 4.3 will be added in Phase 1; items 4.4, 4.5 and 4.6 will be reviewed for inclusion in Phase 2:

4.2 Archival reference and a short summary of the letter. For example, 'BL MS Lansdowne 155 f. 306v - 307v. Instructions for William Herle from the Queen.'

4.3 Observations concerning any unusual features of the letter's physical appearance or copying. e.g. If it is possible to say who the copyist is, i.e. Laurence Johnson, Robert Beale etc. If the letter paper is smaller than usual (does this mean I should take an average size of a few letters?). White space which is unusual, i.e. if the letter appears to begin on the (verso), this feature will be discussed here, as well as text which runs vertically down the left hand margin.

CELL reference: HRL/2005/05/001. Posted: December 2005. Alison Wiggins and Robyn Adams. <http://www.livesandletters.ac.uk/>

4.4 A list of people and places mentioned in the letter e.g. William Humphrey, Richard Huddilstone; Lambhythe, Antwerp, Theobalds.

4.5 Glosses to difficult terms or concepts in the letter e.g. syntactical oddities, obsolete words, or those with different/opposite meanings than modern ones, i.e. 'jealous' 'doubt'. Paraphrase difficult sentences.

4.6 The letter's place in sequence, if relevant e.g. 'This letter forms part of a group of correspondence sent to Lord Burghley and others during Herle's imprisonment in the Marshalsea in the Spring of 1571.' or 'This is one of many letters sent from Antwerp back to England in the summer of 1582 following the attempted assassination of William of Orange.'

5. Address Leaf <address_leaf>

5.1 The address leaf is the verso of the document or outside of the letter packet that typically includes instructions for delivery of the letter. If a Herle letter has an address leaf it is always transcribed and defined with the <address_leaf> element. If a letter has no address leaf, the <address_leaf> element is omitted without comment. The transcribed text must always be contained within the following four child element, that must appear in the order given below. The first three of these elements (superscription, sign_manual, endorsement) can only be nested within <address_leaf>. They ensure accurate capture of the meaning of the different types of text in relation to its spatial and chronological organization by successive annotations by contemporary writers, administrators and modern archivists:

5.2 Superscription <superscription>

The superscription is the address or direction on the outer leaf or verso of the letter. It is defined with the <superscription> element, which can be used once or not at all.

5.3 Sign manual <sign_manual>

The signature of authorization or 'sign manual' is the autograph signature of the sender on the verso of the letter, usually alongside the superscription. It rarely occurs in the Herle letter but, where it does, it is defined with the <sign_manual> element. This element can be used once or not at all.

5.4 Endorsements <endorsement>

The term 'endorsement' is here used to refer to contemporary comments of other kinds than the two above types. It might be a signature, memorandum or remark and is defined with the <endorsement> element. Alan Stewart and Heather Wolfe define 'endorsement' as "a signature, memorandum, or remark usually written on the back of a document" and 'endorse' as "to inscribe (a document) on the back *with* words indicating the nature of its contents, one's opinion of its value, some extension or limitation of its provisions, etc.", (*Letterwriting in Renaissance England* (Washington, Seattle & London: Folger Shakespeare Library, 2004)), p.206. In the case of the Herle letters, endorsements can vary between a note which confirms the direction of the letter, i.e. "from W Herle", usually written by the recipient or recipient's secretary, or a note by the letter-writer himself, e.g. 'Copi of mi discowrse to the Queen'. This often occurs if the document is a copy. In this case it is not clear whether the note on the address leaf is for the letter-writer's own administrative purposes, or whether it is a note to his correspondent clearly defining the document as a copy of a letter, distinct from the original. The presence of these letters in archives deriving from the offices of his patrons suggests the latter. For this reason, the element <endorsement> can be used, as it suggests a note by Herle, but for the use of the recipient. As a

CELL reference: HRL/2005/05/001. Posted: December 2005. Alison Wiggins and Robyn Adams. <http://www.livesandletters.ac.uk/>

rule, any comment written contemporaneously in connection with the letter and its issue can be defined as an endorsement. Any extraneous writing unconnected with the above, i.e. unrelated scribbles or doodles or comments by later, modern archivists will be referred to with the <note type=> element. Where the writer of the endorsement is known, their name or role is recorded as the value of the optional 'resp' (i.e. 'responsibility') attribute. For example:

<endorsement resp="Burghley's secretary">W Herle to my L Mr Wade</endorsement>

The <endorsement> element can be used or any number of times or not at all.

On the address leaf, if 'William' is contracted to 'Wm', by the endorsing hand, I have expanded it to 'William'.

5.5 Later additions <note type="later addition">

Comments of any kind that are by a later hand or hands, i.e. that cannot be associated with the letter's earliest production and reception environments, will be defined with the <note> element. These are, typically, notes by modern archivists, but might also include notes by any owners since seventeenth century. The optional 'type', 'place' and 'resp' attributes can be included with this element where additional information is available or judged to be of particular interest. Where more than one of these attributes is included, their values should not be repetitions: 'type' should always have the value 'later addition'; 'resp' should only be used if the identify of the writer is known; 'place' should only be used if the location of the note is of particular interest or revealing in some way. For example:

<note type="later addition" resp="20th-century archivist">9 nouembr 1574</note>

The <note> element can be used or any number of times or not at all.

6. Letter Text <letter_text>

6.1 The text of the letter proper is defined with the <letter_text> element, which must occur once and only once. The <letter_text> element can contain character data only, or character data and child elements. The following five child elements define features of the letter transcript. They can occur in any order. The first four of these elements (paragraph, list, postscript, subscription, marginalia) can only appear nested within <letter_text>.

6.2 Paragraphs <p>

Here only occasionally writes in paragraphs. The paragraph element <p> is therefore only used where Herle clearly divides the text into identifiable and distinctive 'units'. Where this is not the case, and the division should be described as a mid-passage break rather than a distinctive paragraph unit, the empty element <lb /> ('line break') is used (see 7.4 below).

6.3 Postscripts <postscript>

A postscript is classified as any piece of writing which is evidence of an afterthought. Typically, it is identifiable as it occurs after the signature, or declares itself to be a postscript, or its status as an afterthought is evident from its content and relation to the preceding letter text. Each <postscript> element defines a single postscript. It may be used once, not at all, or many times, depending on the number of postscripts within the letter. An optional 'place' attribute can be included if it is judged necessary to indicate any special features of positioning.

6.4 Subscription <subscription>

The point where Herle subscribes his letters typically occurs after a gradual build up of valedictory forms and it not always possible to locate a single, definite, identifiable point at which this valedictory preparation ends and the subscription proper begins. Nevertheless, the subscription can always be approximately identified and, in the vast majority of cases, its point

CELL reference: HRL/2005/05/001. Posted: December 2005. Alison Wiggins and Robyn Adams. <http://www.livesandletters.ac.uk/>

commencement can be pinned to a specific word or recognizable point in the concluding formula. The subscription text will be defined according to the editor's judgement and using the <subscription> element. For example:

Coronell Steward sent me word by his man yt he had wrytten to yow, & ynclosed those artycles of ye masse within your L lre with som other remebracs, & yett I wolld nott ffayll to send yow on myne allso, <subscription>wherwith I humbly take mi leve, Andwrpe the xith of marche 1581 your L with his harte & shrvice, W Herlleli</subscription>

6.5 Marginalia and Annotations <marginalia>

Herle's marginalia, or the marginalia which occurs on Herle's letters, can be divided into three categories.

Type 1: Marginal notes which are additions to the text made at the time of writing by the person who composed the letter. That is, sentences or longer groups of words that would not fit or would be too cumbersome to be inserted along the line of text as a supralinear insertion. This is regarded as a simple addition to the main text and part of the process of composition. This text will be transcribed at the point it was intended to appear in the text and will be distinguished as an addition by a caret mark at the beginning and end using the caret mark entity ^ (see 7.1 below), so: ^inserted text^. Where Herle or other writer has not indicated a specific place in the text with a caret or other mark, the marginalia tag will be placed after the first word of the proximal sentence. If the word breaks over two lines, then the tag is placed after the word.

Type 2: Marginal notes made as a summary: either one-word summaries or sentences which refer to or continue the proximal paragraph. These are often made by Herle and at least once by Burghley. Where these comments or summaries refer to a particular point in the text they will be transcribed at that point (or, if they refer to a particular word, sentence or paragraph they will be transcribed at the end of that word, sentence or paragraph) and defined with the <marginalia> element. This element has two option attributes 'place' and 'resp'. This 'resp' attribute is mandatory: if the identity or role of the writer is unknown, the attribute is 'unknown hand'. If the writer is known, the value should contain their name in a standard form. The 'place' attribute should be used to indicate the position of the marginalia on the page. As Herle's marginalia and annotations occur consistently on the left-hand margin, the place attribute will only be used if the marginalia deviates from this format. In certain cases where the marginalia seems to be a running commentary on a paragraph's contents, the marginalia will be transcribed in relation to the place where the first marginal annotation occurs (eg. SP 70 127/36). If the paragraph breaks over a page, the marginalia will start again with the new page.

Type 3: Catchwords. Catchwords by a later owner, librarian or archivist will be transcribed and encoded in the same way as all later additions, using the <note type="catchword"> element (see 6.7 below).

6.6 Later additions <note>

Comments of any kind that are by a later hand or hands, i.e. that cannot be associated with the letter's earliest production and reception environments, will be defined with the <note> element. These are, typically, notes by modern archivists, but might also include notes by any owners since seventeenth century. The optional 'type', 'place' and 'resp' attributes can be included with this element where additional information is available or judged to be of particular interest. Where more than one of these attributes is included, their values should not be repetitions: 'resp' should

CELL reference: HRL/2005/05/001. Posted: December 2005. Alison Wiggins and Robyn Adams. <http://www.livesandletters.ac.uk/>

only be used if the identify of the writer is known; 'place' should only be used if the location of the note is of particular interest or revealing in some way. For example:

```
<note type="catchword" resp="archivist">this is the catchword</note>
```

The <note> element can be used or any number of times or not at all.

7. Other features of the encoding (child elements of <address_leaf> and <letter_text>)

7.1 Folio references <folio_ref />

The start of each new folio is indicated with the empty element <folio_ref />. This empty element must be the first child element to appear within <address_leaf>. Within <letter_text>, it must appear at the point where each new folio starts. The attribute 'name' must be included as mandatory and its value is the folio or page reference. For example: <folio_ref name="fol. 16r" /> or <folio_ref name="p. 12" />

7.2 Special characters

Special characters, that is, those other than the standard Latin alphabet and most basic keyboard characters, must be encoded using the appropriate character entity codes, as listed below and according to Unicode standards. The transcripts will be in UTF-8. Any characters that would require UTF-16 will be defined in the project DTD and given an editorial description, rather than using the symbol itself.

Character entity code: *Description of the character:*

&	& [Ampersand]
'	' [Apostrophe]
*	* [Asterisk]
^	^ [Caret]
&crowns;	Crowns, the symbol denoting the monetary unit 'crowns'
†	† [Dagger]
—	-- [Emdash]
–	- [Endash]
>	> [Greater than sign, i.e. when not being used to denote a tag]
<	< [Less than sign, i.e. when not being used to denote a tag]
"	" [Quotation mark]
þ	_ [Thorn, lower-case]
Þ	_ [Thorn, upper-case]

Legal entity codes for any additional special characters that are encountered during the process of transcription (e.g. Greek letters, mathematical symbols, non-standard punctuation) can be found at:

<http://www.webstandards.org/learn/reference/entities.html>

These will need to be added to the herle DTD if the XML is to validate. You will therefore need to keep a record of any new character entities so I can add them to the DTD.

Where Herle has used a caret to denote that a marginal comment relates to this part of the text, I have only encoded the mark if the caret is then included in the marginal annotation as well, see Cotton Titus BVII ff. 44-5.

7.3 Line breaks <lb />

Line breaks will not be noted, unless it Herle has deliberately started a new line in order to indicate a break in the flow of the text. Such breaks will be marked with an empty <lb /> element ('line break'). In the few cases where a letter has been written in paragraphs these will be defined with the <p> element (see 6.2 above).

7.4 Deletions

Where words are deleted but legible, they will be transcribed in the position they appear in the text and defined with the element.. Where there are words deleted and illegible, these will be left untagged, but recorded in an editorial comment; for example: [3 words heavily expunged] or [1 word obscured by ink blot: the?].

7.5 Underlining

Underlining will be represented, although inferences as to who has underlined the document will belong in the Editorial commentary.

7.6 Text that has been emphasized in some way other than underlining

Elaboration such as enlarged, bold or coloured text is very rare in the corpus of Herle letters. Instances that do occur are relevant to understanding stress and emphasis in Herle's often persuasive language. These will be defined with the ('emphasis') element. In cases where Herle has changed script where he writes a word or phrase in a language different to the surrounding text, the attribute and value `type="foreign"` will be used; that is, <em type="foreign">foreign words. Where he has written in a language different to the surrounding text but the script is unchanged, the tag will not be used.

7.7 Lists <list><item>

Herle occasionally includes lists within his letters. The entire list will be defined with the <list> element. Each individual item within the list will be defined with the <item> child element. For example:

```
<list>
<item>This is the first item in the list</item>
<item>This is the second item in the list</item>
<item>This is the third item in the list</item>
</list>
```

The <list><item> elements must always be used in association with one another and they must only be used to encode actual lists (and must not, for example, be used to represent spatial layout or a series of jottings that happen to be written in a narrow column). The text must be transcribed exactly as it appears, with or without preceding numbers or symbols before each item.

8. Other features of the transcription

8.1 Spelling and orthography

Transcriptions will follow the original spelling apart from the modernisation of 'u' to 'v' and 'i' to 'j'. Words beginning 'ff' or 'ff' will be regularised to 'F' or 'f'. In Latin phrases, 'i' has been retained rather than amend to 'j'.

8.2 Contractions and abbreviations

Abbreviations that will be silently expanded include (contraction type give first, then the expansion):

advertismt (advertisement)
ambassadr (ambassador)
governmt (government)
comandmt (commandment)
apprtayne (appertayne)
shrvyce (servyce)
obshrvatyo (observatyon)
vppo (uppon)
sattrday (satterday)
mrvaylous (marvaylous)
mrchant (merchant)
lre/s (lettre/s)
huble (humble)
Sr (Sir)
Ma[ie, tye, ty, ties, tyes ts] (Majestie)
Mr (Master)
Rx/d (Receive/d)
dl/d (deliver/d)
yor (your)
y^e (the)
y^t (that)
w^{ch} (wich)
wth (with)

Abbreviations to be kept will include:

B (Bishop)
P (Prince)
Sp (Spain)
K (King)
D (Duke)
C (Count)
E (Earl)
Q (Queen)
H (Honorable)
R (Right)

Names should be expanded where they are contracted, for example: Tho: (Thomas), Ch. (Charles), Xpofer (Christopher). Where a correspondent refers to 'Wm' (as 'William') this should also be expanded. However, where Herle signs himself 'Wm', the contraction should be kept.

8.3 Punctuation

Punctuation will follow the original. Herle uses full stop, colon, semi colon, apostrophe, quotation marks, forward slash, brackets, as well as the symbols listed above in section 7.2. Where Herle ends his sentences in a forward slash this is transcribed as it appears; it is not indicative of a line ending, although sometimes his lines end in this way by default.

8.4 Curious spatial anomalies

CELL reference: HRL/2005/05/001. Posted: December 2005. Alison Wiggins and Robyn Adams. <http://www.livesandletters.ac.uk/>

Herle's letters uniformly begin at the top-left of the folio page, leaving a gap of about an inch. Therefore, only if a letter does not begin in this way is it necessary to denote the spatial organisation of a letter. Any unusual use of layout or white space will be recorded in the Editorial Commentary (see 4.2 above).

8.5 Changes of hand

Where there is a discernible change in hand, this will be noted in the Editorial Commentary (see 4 above) but no indication will be given within the letter text itself.

8.6 Words omitted in error or haste

Where a word or letter seems to have been omitted and this makes the meaning of a word or passage obscure or difficult to make out, the letter or word will be added by the editor enclosed in square brackets; for example f[rom]. This will include words omitted as a result of error or haste by the author/scribe. It will also include instances where words are gone due to part of the page having been removed. For example, a number of letters in the Cotton collection were singed at the edges in the famous fire; wherever possible in these cases the missing words should be substituted in square brackets from the copy/holograph and a record of the cause of the damage and the copy used will be made in the Editorial Commentary (see 4 above). Where the words are entirely gone and cannot be reconstructed, an ellipsis in square brackets will be given: [...].

9. Use consistent attribute values

It is important that the editorial comments that appear in the text (as the values of attributes, or in square brackets within the transcription [see 7.4 above], or within the Editorial Commentary [see 4 above]) follow a consistent format. They should be clear, accurate, self explanatory, grammatically correct and written in full rather than in abbreviated note form. This is particularly important in the case of attribute values as these will ultimately be output in an automated way (which tends to increase the visibility of any irregularities). Below is an overview of attributes and values, with sample formats.

``

Optional attribute: `type="foreign"`. Always has this value.

`<folio_ref>`

Mandatory attribute: `name="fol. 12r" />`. The value is the folio or page reference and should always be in this format, i.e. 'fol.' or 'p.', followed by a space, followed by the folio or page number, then 'r' or 'v' if necessary.

`<herle_letter>`

Mandatory attribute: `id="1"`. The value is always the id number from the letters catalogue.

`<note>`

Optional attribute: `place="left-hand margin"`. The value denotes the position on the page.

Optional attribute: `resp="Herle"`. The value is the name of the person responsible for writing the note. The name should be in standard form and should always be recorded in the same way in these values. It is a less well known person, the full name should be given.

Optional attribute: `type="catchword"`. The value describes the type of note; for example, catchword, later addition, archivist's comment.

`<postscript>`

CELL reference: HRL/2005/05/001. Posted: December 2005. Alison Wiggins and Robyn Adams. <http://www.livesandletters.ac.uk/>

Optional attribute: `place="left-hand margin"`. The value denotes the position on the page. This attribute should only be included within the `<postscript>` element if the postscript text is positioned in an unusual way, such as lower on the page.

`<subscription>`

Optional attribute: `place="left-hand margin"`. The value denotes the position on the page. This attribute should only be included within the `<postscript>` element if the postscript text is positioned in an unusual way, such as lower on the page.

APPENDIX:

Summary of elements and attributes:

Root:

`<herle>`

Parent element level 1:

`<herle_letter id="id number">`

Parent elements level 2:

`<editorial>`, `<address_leaf>`, `<letter_text>`

Child of `<address_leaf>` and `<letter_text>`:

`<folio_ref name="folio reference" />`, `<note type="later addition" place="where" resp="who by">`

Children of `<address_leaf>` only:

`<superscription>`, `<sign_manual>`, `<endorsement>`

Children of `<letter_text>` only:

`<p>`, `<subscription>`, `<postscript>`, `<marginalia place="where" resp="who by">`

Children of `<address_leaf>`, `<letter_text>`, and of their child elements:

``, ``, `<em type="foreign">`, `<lb />`, `<list><item>`

Alphabetical list of elements and their attributes:

`<address_leaf>`

``

`<editorial>`

`` (optional attribute: type, with value 'foreign')

`<endorsement>`

`<fol_ref />` (mandatory attribute: n)

`<herle>`

`<herle_letter>` (mandatory attribute: id)

`<item>`

CELL reference: HRL/2005/05/001. Posted: December 2005. Alison Wiggins and Robyn Adams. <http://www.livesandletters.ac.uk/>

<list>
<lb />
<letter_text>
<marginalia>
<note> (optional attributes: place, resp)
<p>
<postscript> (optional attribute: place)
<sign_manual>
<subscription> (optional attribute: place)
<superscription>
