



Transcription Standards

These standards are designed to ensure the production of accurate and consistent transcriptions of manuscript sources. They aim to enable the preservation of as much information as possible about the original manuscript in an electronic version encoded in XML. The standards are based on a model that preserves the form and content of the original in a base transcription, but allows an editor to record additional information in an editorial layer, which may be used in the presentation or analysis of the text. Transcribers and editors are expected to preserve a strict separation between the base transcription and the editorial layer, so that the ability to separate the two is maintained. It must always be possible to recover the base transcription.¹

Rules for creating the base transcription

- Characters
 - Original spelling is retained with no modifications for modern conventions of capitalization or spelling.
 - Original punctuation is retained.
 - Superscription of characters is recorded.
 - Modifications to individual characters or groups of characters by lines through or above etc. are recorded.
 - Roman numerals where they occur are recorded as written.
 - Brevigraphs, scribal contractions, scientific symbols etc. are recorded as glyphs (not as descriptions).²
 - Intralineal space is reproduced using the appropriate number of hardspaces.³
 - Hyphenation and catchwords are recorded as they appear.
- Abbreviations
 - Are recorded as they appear and are not expanded in the base transcription.
 - *Expansions may be provided in the editorial layer.*

¹ In CELL we use a software package called Transcriber's Workbench (TW), which is designed to facilitate the implementation of these standards.

² Implementation of the standard requires a method of recording characters and symbols that are used in the manuscript that are not available in the encoding scheme being used by the XML. TW addresses this issue by providing transcribers with glyphs for non-standard characters, scientific symbols etc.

³ XML will ignore whitespace during processing, so gaps recorded using spaces or tabs will not appear in a processed transcription.

- Text layout
 - Line endings are preserved.
 - Blank lines are retained.
 - Page breaks are recorded.
 - A block of text is marked in the base transcription to record its physical position, the direction of the text and whether it is boxed.
 - Where text is divided into columns, the extent of the entire block and the content of individual columns is recorded.
 - *The structural or syntactical significance of a block of text may be recorded in the editorial layer.*
- Deletions
 - If deleted text can be read, the text is recorded and marked as deleted.
 - If deleted text cannot be read, the transcriber records the approximate number of characters using '.' and marks these as deleted.
- Insertions
 - Insertion marks are recorded as characters where they appear.
 - Inserted text is recorded where it occurs.
 - *The linking of inserted text to a specific insertion mark is part of the editorial layer.*
- Missing text
 - Where text is lost through blotting, tearing or other damage to the original, the extent of the missing text is indicated using '.' and marked as lost.
 - *Conjectural readings for lost text must not be included in the base transcription, but may be provided in the editorial layer..*
- Text Decoration
 - Underlining is recorded.
 - The drawing of a box around text within a line is recorded as text decoration.
 - The use of italic or other distinctive forms of script are noted.
 - *Changes of hand are recorded in the editorial layer.*
- Language
 - Text in a language other than that of the main text is marked.
 - *Translations for foreign text may be provided in the editorial layer.*