5 File format basics

5.1 High-level structure

The file is an ASCII text file of unspecified width. There is no fixed line length. The only restriction is that no line shall be wider than 2048 characters[6].

The file starts with a file header, followed by any number of lines.

The file header consists of a single line that has at least 12 characters.

Each of the lines after the file header is of one of two types:

- Comment line
- Data line

Comment lines may appear almost anywhere in the file, and are recognised by a # character in the first position.

The data lines consist of a number of blocks. Each block provides the transliteration of one page, and is organised as follows:

- A single page header
- A number of transliteration items for this page

Every transliteration item consists of:

- A locus identifier
- The complete transliterated text for this locus

5.2 Interlinear files

Interlinear files differ from 'standard' transliteration files, in that many or all transliteration items are repeated several times. Each instance represents a proposed transliteration from a different source. This source is indicated by a transcriber ID[7] that is include in the locus identifier. It is allowed for interlinear files to have only one transcriber ID throughout.

Only one interlinear file has been in extensive use, namely the LSI file identified in Table 2, but this is not available in the IVTFF format. At the time of issue of document version 1.7, no interlinear files in IVTFF format have been published, but this is likely to change.

5.3 Page header (in brief)

A page header, in its most simplified form, has the format:

[6] This number should be assumed as limit by software processing IVTFF files

[7] Terminology clearly distinguishes between transcription and transliteration, and the latter applies to the Voynich MS. However, the word 'transliterator' does not appear to exist, so transcriber is used in this and related documents.