IVTFF – Intermediate Voynich MS Transliteration File Format

R.Zandbergen

File format version 1.7.

Document issue 1.7, 10/04/2020

Contents

1 Purpose

1.1 General

This document describes a file format that allows representation of all existing ('historical') and publicly available digital Voynich MS transliterations in a consistent manner. As the word 'intermediate' suggests, this format is not intended to be the last word, but it will allow the definition of a more permanent format, ideally following accepted standards (such as, for example, 'TEI' [1]), and allow automated machine conversion of all existing transliterations into this permanent format.

Having a standard format for all files also allows the creation of standardised tools and scripts to access all transliteration files.

I would use the occasion to recommend all owners of private (not publicly available) transliterations to consider moving to this format, as it will allow them to use standardised tools, possibly developed by a community of users.

1.2 Limitations

See Section 7.

1.3 Document change history

Issue	Date	Comment
-	(several)	Versions earlier than 1.4 were development versions which have not been released.
1.4	12/08/2017	Changes affecting the format: 1. The symbol used for continuation lines has been changed from \ to / 2. The symbol used to separate alternative readings has been changed from | to : 3. The format version has been added as an optional field to the file header All other changes are of editorial nature. This is the first released version of the format.
1.5	23/08/2017	Mistakes in Table 9 corrected (still had some references to | symbol)
1.5.1	23/09/2017	Additional mistakes corrected (line continuation symbol was still written as \ (instead of /) in Sections 5.5 and 6.2. New rule about version numbering
1.6	21/11/2019	Introduction of the dedicated comment <%> for start of paragraph. Move

[1] "Text Encoding Initiative", see for example http://www.tei-c.org/index.xml . This is only intended to highlight one option.

		to the correct terminology for transliteration instead of transliteration.
1.6.1	19/12/2019	No changes to the format, but add more pre-defined page variables and include an overview in Annex 2.
1.7	10/04/2020	Introduction of text tags similar to page variables. Further clarification and updates related to interlinear files. Annex 2 moved to [R-2]

1.4 References

[R-2] IVTFF – Conventions (to be issued)

[R-3] ivtt tool user manual, issue 1.1 of 10/04/2020. Available via: http://www.voynich.nu/software/ivtt/IVTT_manual.pdf

[R-4] Web page: http://www.voynich.nu/transcr.html

[R-5] Web site: http://www.voynich.nu/

2 Definitions

Table 1 lists the most important terms used in this document, and explains their meaning. Any text in the explanation that has been written in blue italics is a reference to another term in the same Table.

Table 1: List of terms used in this document

Term	Meaning
MS	The Voynich Manuscript, a codex that consists of a blank cover, paper end leaves, and parchment ***leaves***. It includes text and illustrations. The vast majority of the text is written in an otherwise unknown alphabetic script. There are also a few entries in the normal (Latin) alphabet.
Voynichese	The name used here for the unknown script in the ***MS***.
Text block	The complete set of parchment ***leaves*** of the ***MS*** that have been bound together.
Quire	(Also: gathering). One of the 18 sets of stacked parchment ***sheets*** that together form the ***text block***.
Bifolio (also: sheet)	A large piece of parchment folded (usually) in the middle and bound into the ***MS***. A complete ***bifolio*** consists of two ***folios***. (There are only 2 incomplete ***bifolios*** in the ***MS***).
Folio (also: leaf)	A double-sided piece of parchment, usually half of a ***bifolio***. The ***folios*** in the MS have been numbered 1-116 though a few are missing. All ***folios*** have text and/or illustrations on both sides. The 'front' of each ***folio*** is called the recto side, and the 'back' is called the verso side.
Foldout	A ***folio*** that is two or more times as wide as a standard ***folio***, and has been folded to fit into the ***MS***.
Page	One side of a ***folio***. ***Foldout folios*** are two or three ***pages*** (or panels) wide, and, depending on the layout, may have several separate *pages* on each side.
Foliation	Handwritten numbers in the corners of every ***folio***, running from 1 to 116. Several numbers are skipped, as if these ***folios*** have been lost.
MS text	In the present format description document, the ***MS text*** is considered to be all text in the ***Voynichese*** script. Text in the Latin alphabet is not considered.
Eva	One of the conventional transliteration alphabets.
Page name	An identifier of a ***page*** in the *MS*. It consists of the letter f, followed by the ***folio*** nr, followed by r or v (for recto or verso), and optionally by a number (1-3) in case it is a ***foldout folio***. This is described in more detail below (Section 4.2).
File header	The first line of an IVTFF transliteration file
Page header	A piece of text that identifies the start of a new ***page*** in the transliteration file. It is described in more detail below.
Page variable	An annotation included in a ***page header*** that defines a testable property of the entire ***page***.
Locus	An identifier for a unique piece of ***MS text***. It is described in more detail below.

Transliterated text	A string of characters that represents a piece of ***MS text***.
Transliteration item	A ***locus*** identifier followed by some ***transliterated text.***
Interlinear file	A transliteration file that repeats the same ***transliteration item*** several times, on separate lines, presenting transliterations from different sources in parallel
Transcriber ID	A single-character code identifying the transliteration source in an ***interlinear file***
Text tag	An annotation included in a ***transliteration item*** that defines a testable property of the present ***locus*** and the following ***loci***.
<space>	A space character (ASCII code: decimal 32 or hex. 20).
Whitespace	A ***<space>***, or several ***<space>*** characters in succession.

3 Introduction

3.1 General

All historical digital Voynich MS transliterations have been represented in plain ASCII text files, and the IVTFF format is no exception. Following is an overview of these past transliteration efforts[2].

Table 2: List of historical digital transliterations

Code	Transcriber(s)	Date	Comment
FSG	William Friedman and his 'First Study Group'	1946?	This is based on a unique, dedicated transliteration alphabet ('FSG')
C-D	Prescott Currier and Mary D'Imperio	1970's	Using a new transliteration alphabet ('Currier'). The resulting transliteration is also known as 'voynich.orig'
Vnow	Reeds, Gillogly, Guy (and others)	1990's	Update of the C-D transliteration made by members of the Reeds/Gillogly mailing list (using the 'Currier' transliteration alphabet). The resulting transliteration is also known as 'voynich.now'
TT	Takeshi Takahashi	Nov. 1998	The first almost complete transliteration of the MS, using the 'Eva' alphabet.
IT	Takeshi Takahashi	Nov. 1998	Copy of Takeshi's transliteration that has been included in the interlinear file (LSI, see below). It is slightly modified from the original.
LSI	(Many)	1999	A file prepared by Gabriel Landini based on efforts by Jim Reeds, containing several historical transliterations in an interlinear manner. It was significantly updated by Jorge Stolfi, so we may refer to it as the Landini-Stolfi Interlinear (LSI) file.
LZ	Landini and Zandbergen	1999	Two independent, complete transliterations based on agreed rules and conventions. The intention was to merge them but this was not completed. The result has not been published.
GC	Glen Claston	Early 2000's	A new transliteration of the entire MS made by a single person, based on a new transliteration alphabet.
ZL	Zandbergen	(on-going)	A published version of Zandbergen's part of the LZ transliteration, updated to cover the complete MS. It is (still) undergoing final verification.

The coverage of the text in the MS that is represented in these files has increased over time, and the file indicated as "ZL" is the first published file that covers the complete MS.

[2] For more information about these files, see [R-4].

3.2 Purpose

The purpose of transliteration files is two-fold. On the one hand they can be read and interpreted by humans. For this purpose, numerous types of annotations and formatting are included in a human-readable manner. Certain layout conventions (use of whitespace, continuation lines, empty lines) are included to make the text more easy to read.

On the other hand, the files are meant for machine processing. In this case, the above conventions are not useful. In some cases, alternative representations of items are possible, which favour one or the other representation. These will generally be indicated in this document.

4 Structure of the MS

4.1 General

The Voynich MS is a book or codex which is composed of parchment leaves or folios, combined into gatherings or quires[3]. A 'standard' quire in the MS consists of a stack of four sheets or bifolios, which is folded in the middle to form 8 folios. Each quire is sewn onto a set of three thongs, at the combined fold. A standard Voynich MS bifolio is roughly 32 cm wide and 23 cm high, meaning that a folio is roughly 23 by 16 cm.

All folios in the MS have writing and/or illustrations on both sides, and the individual sides of each folio will be referred to here as pages. Thus, a standard quire has 4 bifolios, 8 folios or 16 pages.

The quires in the Voynich MS have been numbered 1 to 20 (with 16 and 18 missing). Like many other medieval codices, several quires in the Voynich MS do not consist of the standard 8 folios. What is more unusual, though, is that several bifolios are wider than the standard size. These have additional folds and consequently more than the normal four pages. They are referred to as foldouts. These foldouts have different dimensions, with widths of the corresponding bifolios ranging from three to five pages (instead of two). In addition, there is one (approximately) 45 by 45 cm bifolio which has an additional horizontal fold.

Every folio in the MS has a folio number written on it, with the numbering running from 1 to 116, though 14 folios are missing. The corresponding folio numbers are skipped.

4.2 Page naming (numbering)

The notation used to identify a page in the Voynich MS is the character f (for folio) followed by the folio number, followed by r (for recto - the front) or v (for verso - the reverse). Thus, the first quire starts with pages: f1r, f1v, f2r, f2v, f3r, etc, and ends with f7v, f8r, f8v. The four pages: f1r, f1v, f8r and f8v together form one bifolio.

For the foldout folios, the following additional rule is applied. When any foldout is completely folded out, to the right of the binding gutter one sees the recto side of this folio. To the left is the verso side of the previous folio. If the folio nr. is *n*, the recto 'pages' of the foldout are numbered left to right (i.e. away from the binding): f*n*r1, f*n*r2, etc. On the verso side (with the binding to the right) the 'pages' are numbered right to left (again away from the binding): f*n*v1, f*n*v2, ...

This is illustrated below for the single bifolio in Quire 11, where the red mark represents the binding. To its left we see the standard folio f71 and to its right the multiple foldout folio f72:

[3] This text is from http://www.voynich.nu/descr.html

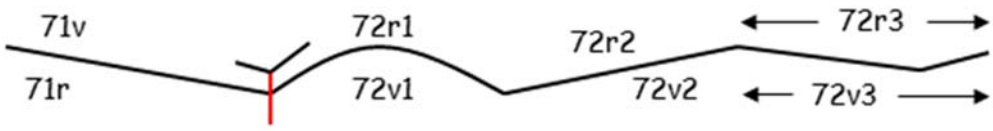


Figure 1: page definition for folio 72 (example)

The creases of the foldout folios sometimes form boundaries between distinct pages and sometimes do not. That is, continuous lines of writing sometimes (though rarely) cross foldout creases.

A special case is quire 14 with the pair of folios f85 and f86, which form the above-mentioned multiple-foldout with the horizontal folding crease. Beside the horizontal fold it also has two vertical creases, so the bifolio is divided into 6 panels: 3 above the horizontal crease and 3 below. Each panel has approximately the same size as a normal 'page'.

When the sheet is folded out, a single, complicated drawing is visible, which covers six panels and is usually referred to as the Rosettes page. This single drawing covers the verso side of f85 and the recto side of f86, and the page name that will be used for this is 'Ros'. The list of pages for this quire is therefore:

- f85 recto, with two pages: f85r1 and f85r2
- f85 verso + f86 recto, with one single drawing. One page: fRos
- f86 verso, with four pages: f86v4, f86v3, f86v6, f86v5

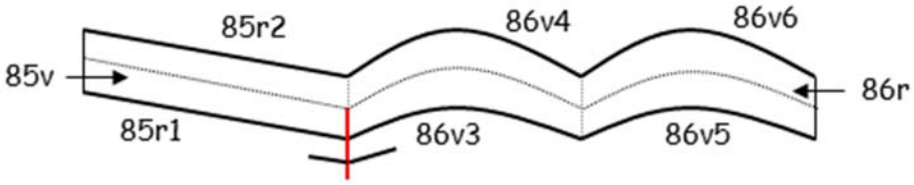


Figure 2: Layout of quire 14 and its folios and pages

4.3 Structure by illustration type

Almost all pages in the MS are illustrated. Illustrations of a similar type are mostly grouped together in the MS, while there are also a few text-only pages among them[4]. The following list indicates the different types of illustrations that may be found in the MS:

- Herbal or botanical, with drawings of herbs, some of which look realistic, while others appear imaginary
- Astronomical, with illustrations of Sun, Moon, stars and zodiac symbols
- Cosmological, with mostly circular drawings
- Biological or balneological, with some possibly anatomical drawings with small human (mostly feminine) figures populating systems of tubes transporting liquids

[4] Text summarised from http://www.voynich.nu/descr.html

- Pharmaceutical section, so called because it has drawings of containers, next to which various small parts of herbs (leaves, roots) have been aligned
- Marginal stars, in a section that contains over 300 short paragraphs

4.4 Types of text items

The text of the MS has been written mostly in a line-by-line manner, obviously from top to bottom and from left to right[5]. The majority of this text is written in short paragraphs, which are often separated from each other by a somewhat larger line spacing. The text tends to have a straight left margin, and is mostly only roughly right-justified, except for the last line of each paragraph which tends to be shorter. Occasionally this last line is centred or right-justified, and also occasionally, the shorter last line is left-justified, but has an additional word in the right margin, a so-called 'title'.

The text consists of groups of characters separated by spaces, and these groups seem to form words.

In some places, single 'words' are written near elements of drawings. These have come to be called 'labels'. There are also places in the MS, for example in the cosmological section, where single words appear as elements in the overall design, but not necessarily near an identifiable object.

A number of pages have circular drawings, some with text written in normal paragraphs, but all with text that has been integrated in the drawings. Frequently, text is written along the circumference of these circles, and occasionally also along radial lines.

[5] Text summarised from http://www.voynich.nu/writing.html

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.

<*page-name*>

Where the left caret has to be in the first position of the line, *page-name* is the page name as identified in Section 4.2, and there are no <space> characters between the carets.

5.4 Locus identifier (in brief)

A locus identifier, in its most simplified form, has the format:

< *page-name . number , type*>

Where the left caret has to be in the first position of the line, *page-name* is the page name as identified in Section 4.2, *number* is a number increasing from 1 for each page, and *type* is a code explained further below (see Section 6.4).

5.5 Transliterated text (in brief)

The transliterated text for each locus is on the same line as the locus identifier.

The file format does not prescribe which transliteration alphabet shall be used to represent the transliterated text. By default, any file shall use only one single alphabet. This alphabet is defined in the file header (see Section 6.1).

A number of characters in the transliterated text have a special meaning. No transliteration alphabet is allowed to use these characters. These special characters are listed in Table 3, and have been chosen such, that they do not clash with any transliteration alphabet known to me (for which see Table 4 below). The special meaning of each of these characters will be described in more detail further below (Section 6.5).

Table 3: list of characters with a special meaning in the transliterated text

Character(s)
< > . , { } [:] @ ; ? /

If a piece of transliterated text is considered too long to be represented on a single line, it can be wrapped to the next line, by terminating the line with a / (slash), and continuing on the next line which must also have a / (slash) in its first position, which should be followed by at least one <space> character before the transliterated text is continued. This feature is specifically included to facilitate human reading of the file.

5.6 Use of whitespace

Whitespace has no meaning, and is used only to improve the visual appearance of the file, again for human reading. Following are rules and recommendations related to the use of whitespace:

- Whitespace is expected (but not obligatory) between the locus identifier and the transliterated text. It is recommended to start the transliterated text in character position 19.

- Whitespace is not allowed inside locus identifiers
- Whitespace should only appear among the transliterated text in interlinear files
- The use of whitespace after the transliterated text, at the end of a line, is deprecated
- One <space> should appear before a / at the end of a line
- One <space> should appear after the / at the start of a continuation line

5.7 Comments

The format allows three types of comments:

- Comment lines (already mentioned above)
- In-line free comments
- In-line dedicated comments

5.7.1 Comment lines

Comment lines may appear almost anywhere, and are recognised by:

- Obligatory # sign in the first position of the line
- Recommended (optional) whitespace following the # sign

Apart from this, any character may appear in comment lines, in any order, without restriction.

5.7.2 In-line free comments

In-line free comments may appear as part of the transliterated text, for example to annotate something. These comments are intended for human interpretation of the file.

An in-line free comment starts with the character pair: <!

It is terminated by the first appearance after this of the character > which has to be on the same line.

It may include any sequence of characters, also whitespace.

5.7.3 In-line dedicated comments

A few in-line dedicated comments are defined, e.g. <-> and <$> . Their meaning is explained further below (Table 10: characters defining different types of in-line comments). These comments are primarily intended for machine processing, but can be equally informative for human interpretation of the file.

5.8 Page variables and text tags

The format foresees two similar methods of annotating the text with meta-data. The first of these, page variables, are valid for an entire page. These variables are defined in the page headers. The second, text tags, are set using dedicated comments in the transliterated text. These values are valid from that line of text onwards, until the end of the page, or until they are overruled by a new setting of the same tag. The format of the two is very similar. Page variables are set using: $X=y while text tags are set using @X=y. The variables and tags share the same meaning.

5.9 Example

The following pseudo-transliteration text is intended to illustrate the basic layout of the IVTFF format:

```
#=IVTFF Eva-  1.4
#
<f1r>
# Start of page 1
#
<f1r.1,@P0>     fachys.ykal.ar<!strange r>.ataiin.shol.shory.cthres.y.kor.sholdy
<f1r.2,+P0>     sory.ckhar.or,y.kair.chtaiin.shar.ase.cthar.cthar,dan
<f1r.3,+P0>     syaiir.sheky.or.ykaiin.shod.cthoary.cthes.daraiin.sy
```

6 File format details

The high-level file structure definition in Section 5.1 fully applies and is not repeated here.

6.1 File header

The file header is the first line in the file. It must have at least 12 characters, but may be longer. The first 8 characters must be the sequence `#=IVTFF` followed by one <space>.

The four characters after this are used to identify the transliteration alphabet that is used in this file.

Following this code may be an indication of the version of the IVTFF format definition. This is of the format `A.B`, as given on the title page of this document. The format definition document may have a version number consisting of three parts (`A.B.x`), which refers to format definition `A.B` regardless of the values of `x`. The format indicators in the file are only meaningful for versions 1.5 or higher.

For the identification of the transliteration alphabet, the following cases have been pre-defined:

Table 4: List of pre-defined transliteration alphabet codes

4-character code	Char	Description
`FSG-`	F	The alphabet agreed by Friedman and his team
`Curr`	C	The alphabet used by Prescott Currier
`FGuy`	G	The 'frogguy' alphabet by Jacques Guy
`Eva-`	E	Eva (either basic or extended Eva)
`v101`	V	The voynich-101 alphabet by Glen Claston

Additional codes may be added by users.

The single-character code is intended to allow the use of several different transliteration alphabets in one file. It can be used in a dedicated in-line comment as described in Table 10. This is not yet used in any transliteration file, and is reserved for a future extension.

6.2 Comment lines

Comment lines are recognised by:

- Obligatory # sign in the first position of the line
- Recommended (optional) single <space> character following the # sign

The comment line terminates at the end of the line.

Any character may appear in comment lines without restriction, and reserved characters are not interpreted. The /character can therefore not be used to continue a comment line. A block of comment lines needs to have a # in the first position of each line.

Comment lines may appear anywhere in the file, except after lines that end with a / character.

The first line of the file (file header) looks like a comment line, and may be interpreted or ignored by tools and scripts.

6.3 Page headers

A page header has the following format:

`<page-name>  <! page-vars>`

Where *page-name* is the page name as described in Section 4.2. A complete list of page names in the order in which they must appear in any transliteration file is provided in Annex 1.

The dedicated comment `<! page-vars >`, which is specific for the page header, is used to set a number of one-character variables to a one-character value, for the entire page, as already described in Section 5.8. For example, the following dedicated comment:

```
<! $A=1   $B=b    $C=C >
```

sets variable A to 1, variable B to b and variable C to C, for the present page. All variable definitions set for a previous page are reset at the start of a new page. Variable names must be upper case characters, but their values can also be lower case or numerical.

In addition, the value "@" may be assigned to some variables, which indicates that the page variable will be overruled by text tags (for which see also Section 5.8). More information about text tags is provided in Section 6.7.

The following variables are used in the transliteration files that are available via [R-4]. See also the annex in [R-2].

Table 5: List of pre-defined page variables (also used for text tags)

Variable	Values	Meaning
$Q	A – T	Quire number, from 1 to 20. P (16) and R (18) are not used.
$P	A – X	Page number within quire. For the possible values, see Annex 1 and [R-2].
$F	a-f , u-z	Folio number within quire. For values see [R-2].
$B	1-6	Bifolio number within quire. Counted outside to inside.
$I	(see below)	Illustration type on this page (see below)
	A	Astronomical (excluding zodiac)
	B	Biological
	C	Cosmological
	H	Herbal

	P	Pharmaceutical
	S	Marginal stars only
	T	Text-only page (no illustrations)
	Z	Zodiac
$K	Y	Has a key-like sequence. (If absent: has none)
$L	A, B	Currier language of this page, or text section
$H	1, 2, 3, 4, 5, X, Z, @	Writing hand used on this page. In files up to format version 1.6, equal to the Currier hand. From 1.7 onwards reserved for the hands identified by Lisa Fagin Davis.
$C	1, 2, 3, 4, 5, X, Z	Currier's hand used on this page
$X	(see below)	Has extraneous writing. (If absent: has none)
	@	As specified by text tags, allowing to identify several cases on one page.
	C	A colour annotation
	M	A month name
	O	Other
	S	A sequence of characters or numbers
	V	Various (a combination of the above) - deprecated
	Y	Any kind (legacy files only)

Additional variables may be added by users.

6.4 Locus identifiers

Locus identifiers have the following format:

`< page . num , code >`

Or :

`< page . num , code ; T >`

Whitespace is not allowed inside locus identifiers, but it is used in the patterns above for clarity.

The fields have the following meaning:

Table 6: Definition of the fields in the locus identifier

Field name	Explanation
page	The page name, which has to match the most recent page header.

num	A sequence number, incrementing from 1 for each page. The highest number that presently occurs is 160.
code	A 3-character code, which is a 1-character 'locator' followed by a 2-character locus type
T	An optional single-character transcriber ID. Only used in interlinear files that include several parallel transliterations.
(locator)	Indication of the relative position of this locus.
(locus type)	Upper case character followed by lower case character or number, indicating which type of text item this locus describes.

Within the scope of the IVTFF format, the values of *num* have been defined for all loci in the MS, and a document (possibly a web page) describing this will soon be published. None of the historical transliterations include all loci.

The preferred order to describe loci in any transliteration file is defined in Section 6.7.

Following is a list of possible 'locator' values.

Table 7: Definition of 'locator' characters

Character	Meaning
@	The position of this locus is unrelated to the previous item, or not easily described by one of the following. This locator is always used for the first item on each page.
+	This locus is generally below the previous item. This is the most common case.
*	The locus is at the start of the line below the previous item, but at the left margin, while the previous item was not.
-	The locus is on the same line as the previous item, but across a drawing element (future extension, not yet used)
=	The locus is on the same line as the previous item, but separated by some white space.
&	Similar to = but along a circular line
~	The same as - or =, but indicating that the vertical alignment is not good.

The locus type (complete type) consists of a generic type (capital letter) followed by a subtype. The valid subtypes depend on the generic type, as defined in the following table.

Table 8: Defintion of locus types

Generic Type	Complete Type	Meaning
P		Linear text in paragraphs
	P0	Normal left-justified text
	P1	Normal paragraph text that is significantly away from the left, typically due to a drawing, or some other text to its left.
	Pb	A free-floating set of lines in a non-standard location.
	Pc	A roughly centred line (compared to the previous line)
	Pr	A (roughly) right-justified line
	Pt	A right-justified 'title' which is on the same line as the previous item
L		Short piece of text, a word, or a character that is anywhere on the page. Mostly these are the so-called labels
	L0	It is not clearly near any drawing element
	La	A label of an astronomical or cosmological drawing element, which is not a star or a zodiac element
	Lc	A label of a container in the pharmaceutical section
	Lf	A label of a fragment of a herb in the pharmaceutical section
	Ln	A label of a nymph in the biological section
	Lp	A label of a large herb or plant drawing (in the herbal section)
	Ls	A label of a star
	Lt	A label of a tube or a tub in the biological section (essentially everything that is not a nymph)
	Lx	Extraneous writing (e.g. in the margin)
	Lz	A label of a zodiac element
C		Text along the circumference of a circle
	Ca	The text runs anti-clockwise
	Cc	The text runs clockwise
R		Text along the radius of a circle
	Ri	The text runs outside to inside (inwards)
	Ro	The text runs inside to outside (outwards)

6.5 Alternative locus identifiers

For machine processing, an alternative type of locus identifier is supported in parallel. This also has a '<' in the first position, but the locus ID consists of only 5 characters. It does not have any locator information but it supports the optional transcriber ID. It terminates with a '>' in position 7 or 9.

The five-character code consists of two upper case characters followed by three digits.

The two characters identify the page using the values of the $Q and $P page variables as defined in Annex 1. The three digits give the value of *num* as per Table 6, with leading zeroes. Thus, the following are valid 'alternative' locus identifiers:

```
<AA001>
```

```
<AA001;Z>
```

If a file uses alternative loci, the page header uses '000' for the value of *num*, i.e. the page header for f1r is:

```
<AA000>
```

In files using these alternative locus identifiers, it is recommended to start the transliterated text in character position 11.

6.6 Transliterated text

The transliterated text comes after each locus identifier, after some optional whitespace, and includes characters of the transliteration alphabet, with, in addition, any of the following special characters:

Table 9: List of special characters in transliterated text

Char.	Meaning
/	If this character appears, it must be the first or the last character in the line. It does not represent a voynichese character, but indicates wrapping of the transliterated text for a locus over two (or more) lines. Comment lines are not allowed between continuation lines. A line following one that ends with a / must also have a / in the first position. The / cannot appear inside brackets of the type [ ] or { } (see below) , and if it appears inside an in-line free comment (<! >) , it loses its special meaning and is just a text element.
.	This character represents an apparent word space in the MS text.
,	This character represents an uncertain apparent word space in the MS text, meaning that the transcriber had doubt that a space between two characters was sufficiently wide to call it a word space.
<	If this character appears as part of the transliterated text, it is not the start of a page header or locus identifier (which must have the < in the first position in the line), but it is the start of an in-line comment. More about in-line comments may be found below this table. Every in-line comment must be closed by a > on the same line in the file. It is permitted to have several < … > pairs on the same line.

>	End of an in-line comment. This character is not allowed to appear unless there was a preceding <
@	Start of a representation of a high-ascii character. It must be followed by 3 digits and then a semi-colon (;). The 3 digits must read a number that is at least 128 and not greater than 255. It is permitted to have many such high-ascii codes on the same line.
;	End of a representation of a high-ascii character. This character is not allowed to appear unless there was a preceding @ followed by three digits.
{	Start of a representation of a ligature of Voynichese characters. The ligature must be closed by a } on the same line. The only characters that are allowed to appear between { and } are valid characters of the transliteration alphabet, and high-ascii characters represented as *@lmn;* It is permitted to have many ligatures on the same line.
}	End of a representation of a ligature. This character is not allowed to appear unless there was a preceding {
[	Start of a representation of an uncertain reading. It must be followed by a] on the same line, which marks the end of the uncertain reading. An uncertain reading means that the transcriber is not certain which character is meant, and gives two or three options separated by colons. It is always understood that the most likely option is the first one in the list. If there are only two options, and each option is represented by a single character in the applicable transliteration alphabet, it is permitted to leave out the colon . It is permitted to have many uncertain readings on the same line.
]	End of a representation of an uncertain reading. This character is not allowed to appear unless there was a preceding [
:	This character only has a special meaning if it appears inside a [ ] pair. It is used to separate the different options for the alternative readings.
?	A single unreadable character.
???	An unknown number of unreadable characters.

There are several types of comments, which are recognised by the character that immediately follows the <

Table 10: characters defining different types of in-line comments

Char.	Meaning
!	This is the start of a free comment. There may be any number of characters between the ! and the > that closes it. The > must be on the same line.
@	Set a text tag. <@X=y> sets text tag X to value y. <@X=@> 'un-sets' it. See also Section 6.7.
-	The sequence <-> means that the text is interrupted by a drawing element
%	The sequence <%> means that this is the start of a paragraph. This may appear anywhere in a line, but should sensibly be placed at the start, possibly only preceded by other comments.

$	The sequence <$> means that this is the end of a paragraph. This should appear at the end of a line.
~	The sequence <~> means that the part on the left and the part on the right are not well aligned vertically. If there is a drawing element in between, it should be written as <-><~>
:	This introduces a transliteration alphabet. There has to be exactly one character after this, and before the closing > . If this character is a 0, it means that the following text is in the default transliteration alphabet of the file, as identified in the page header. If it is any other character, its meaning is defined in Table 4 above. This is not yet used in any transliteration file, and is reserved for a future extension.

6.7 Text tags

Text tags may be specified on any line of transcribed text, using a dedicated comment of the type: <@X=y>. By convention, they should be located at the start of the line. The effect of the text tag is the same as that of page variables, except that they apply only to part of the page, namely starting from the line on which they are given, until the end of the page, or until the tag is set to an alternative value.

The following conventions apply:

1. There is no difference between a page variable and a text tag, apart from the scope where it is defined / set. The possible values of text tags are those listed in Table 5.
2. If text tags are used on a particular page for variable X, the page header for this page should include the variable setting $X=@.
3. If a page variable setting specifies the value @, the value of the variable/tag before the first text tag is 'unset'.
4. If a text tag sets the tag X to value y, this value applies to the entire line on which the tag is set, until the end of the page, or until it is re-defined by a new occurrence in a later line.
5. It is not allowed to set the same tag to two different values on one line, even though tools may ignore this.
6. If a page variable is not set to @ for a particular page, text tags should not be used to modify it. However, tools may ignore these tag settings.

6.8 The order of items in the file

1. It is not mandatory for transliteration files in the IVTFF format to be complete, or even to include all pages of the MS.
2. Those pages that are included in the file must be in the order defined in Annex 1.
3. It is strongly recommended, that the loci for any page are ordered according to the item *num* as part of the locus ID (see Section 6.4).
4. If a particular transliteration file does not have any transcribed text for a particular page, that page header should not exist in the file (but it is not forbidden).
5. If a particular transliteration item does not have any transcribed text, the corresponding locus should not exist in the file (but it is not forbidden).

6.9 Minimum conformance

Tools reading an IVTFF file shall be able to rely on a minimum conformance to the format definition, which is given by the following rules. Files not observing these are not conformant.

- The first line matches the file header definition of Section 6.1, though without restriction on the 4-character transliteration alphabet code
- The first character of each line can only be one of three characters: #, < or /
- If the first character of a line is < , the line is a page header or the start of a new locus
- If the first character is / then the previous line must have ended with a /
- Page headers must have a valid page name between < and > (see Annex 1) and no whitespace
- Locus identifiers have no whitespace, refer exactly to the page in the most recent page header, have an integer number from 1 to 999 between the . and the , and a three-character item between the , and the >
- An optional transcriber ID consists of a single character after a semi-colon
- All in-line free comments <! are closed by a > on the same line in the file
- All other in-line comments consist of 3 or 5 characters, of which the first is < and the last is >
- All alternative readings [are closed by a] on the same line in the file
- All ligature indications { are close by a } on the same line in the file
- All high-Ascii codes starting with @ are immediately followed by three digits giving a number in the range 128 to 255, and the three numbers are followed by a semicolon ";".

7 Limitations

7.1 Introduction

The present format definition ‘evolved’ largely in three stages.

The earliest stage was a set of conventions used in the 1990’s by people communicating through the internet (see also Ref. [R2]). This established some of the most basic usage, such as < > to identify loci, # for full-line comments, and . to indicate a word break in the text.

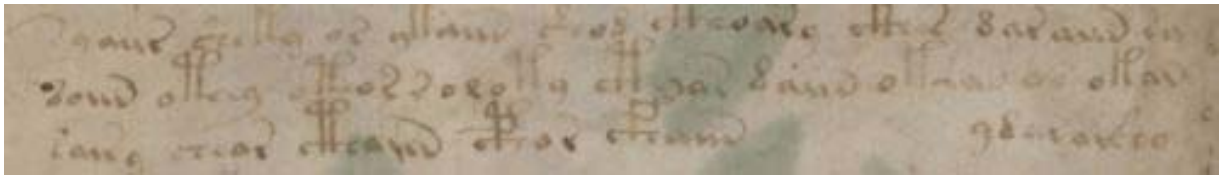
In a second stage, it was extended based on Unix-like notations such as { } for in-line comments, [ ] for alternate readings, ( ) for grouping and &…; for special characters.

In the present, third stage, many of these had to be changed in order to accommodate text in the v101 alphabet which uses many of these characters. As a result, some elements of this format definition are no longer as intuitive as they were in earlier definitions.

7.2 Limitations

1. In case transliteration files are not complete, or the loci are not sorted in increasing order, the ‘locator’ field can be potentially misleading. It is therefore preferred to re-arrange also historical transliteration files according to increasing locus order. It should be always understood that the meaning of the locator is w.r.t. the previous ‘number’, not necessarily the previous entry in the file
2. It has not always been possible to sort the loci in a completely consistent manner. This is most clearly the case in the presence of vertical lists. In some cases, the vertical arrangement appears the dominant one (e.g. on f66r), whereas in others, the horizontal alignment appears dominant (e.g. on f49v).
3. The so-called ‘interlinear placeholders’: ! and % which appeared in the LSI file, clash with the v101 transliteration alphabet. They are no longer used in the IVTFF format. Space characters shall be used in case the user wishes to create interlinear files where characters line up vertically.

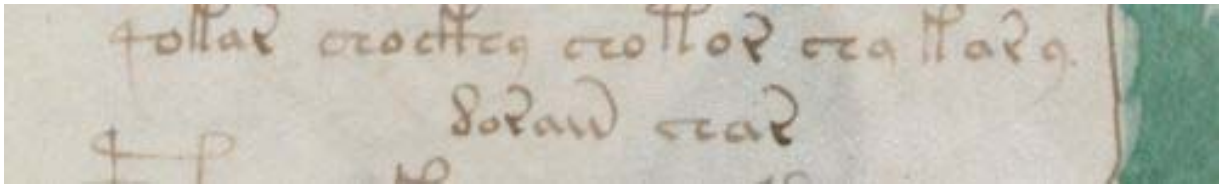
8 Examples



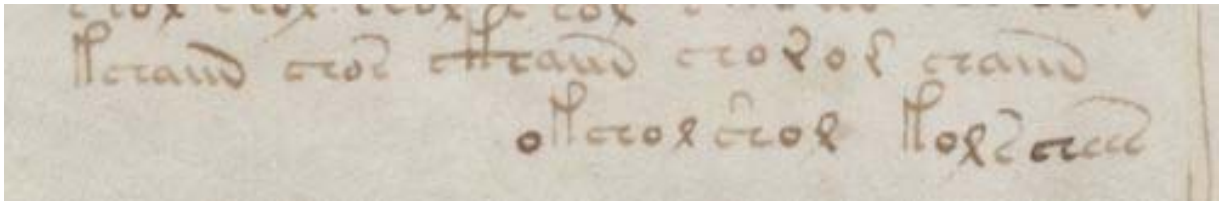
```
<f1r.3,+P0>      <%>syaiir.sheky.or.ykaiin.shod.cthoary.cthes.daraiin.sy
<f1r.4,+P0>      soiin.oteey.oteo[s|r],roloty.cthiar,daiin.okaiin.or.okan
<f1r.5,+P0>      sair,y.chear.cthaiin.cphar.cfhaiin
<f1r.6,=Pt>      ydaraishy<$>
```

```
<f69v.1,@Cc>     dair.cheyky.otaza.sar,ar,chykar.okoirsh,ar.chetody.<!truncated>
<f69v.2,@Cc>     yka{ckh'h}.chol.ykar.dal.ykady.[i|?]okeeor.cheey.choly.<!truncated>
<f69v.3,@Cc>     doair.otaldal.dair.@173;.chdy.otoar.ar,{y'},chy.qoteor.<!truncated>
...
<f69v.5,@Ri>     <!10:00>okeo,dy
<f69v.6,@Ri>     <!10:30>ochoyk
<f69v.7,@Ri>     <!11:00>ykeey
<f69v.8,@Ri>     <!11:30>ytory
<f69v.9,@Ri>     <!00:00>oeesy
```

<f68r1.11,@Ls>	okoaly
<f68r1.12,@Ls>	chocphy
…	
<f68r1.15,@Ls>	otydy
<f68r1.16,@Ls>	okear
<f68r1.17,@Ls>	cphocthy
…	
<f68r1.23,@Ls>	otchdo



<f42r.5,+P0>	qokar.chockhy.chotor.chy.kary
<f42r.6,+Pc>	dorain.char<$>



<f42r.22,+P0>	kchaiin.chos.ckhaiin.choro,r.chaiin
<f42r.23,+Pr>	okchol,shol.kolschees<$>

Annex 1: List of pages in the MS

Quire	Page	$Q $P	Use in file?	Th. Petersen page nr	Comment
1	f1r	A A	Y	1	
1	f1v	A B	Y	2	
1	f2r	A C	Y	3	
1	f2v	A D	Y	4	
1	f3r	A E	Y	5	
1	f3v	A F	Y	6	
1	f4r	A G	Y	7	
1	f4v	A H	Y	8	
1	f5r	A I	Y	9	
1	f5v	A J	Y	10	
1	f6r	A K	Y	11	
1	f6v	A L	Y	12	
1	f7r	A M	Y	13	
1	f7v	A N	Y	14	
1	f8r	A O	Y	15	
1	f8v	A P	Y	16	
2	f9r	B A	Y	17	
2	f9v	B B	Y	18	
2	f10r	B C	Y	19	
2	f10v	B D	Y	20	
2	f11r	B E	Y	21	
2	f11v	B F	Y	22	
2	f13r	B I	Y	23	
2	f13v	B J	Y	24	
2	f14r	B K	Y	25	
2	f14v	B L	Y	26	
2	f15r	B M	Y	27	
2	f15v	B N	Y	28	
2	f16r	B O	Y	29	
2	f16v	B P	Y	30	
3	f17r	C A	Y	31	
3	f17v	C B	Y	32	
3	f18r	C C	Y	33	

3	f18v	C D	Y	34	
3	f19r	C E	Y	35	
3	f19v	C F	Y	36	
3	f20r	C G	Y	37	
3	f20v	C H	Y	38	
3	f21r	C I	Y	39	
3	f21v	C J	Y	40	
3	f22r	C K	Y	41	
3	f22v	C L	Y	42	
3	f23r	C M	Y	43	
3	f23v	C N	Y	44	
3	f24r	C O	Y	45	
3	f24v	C P	Y	46	
4	f25r	D A	Y	47	
4	f25v	D B	Y	48	
4	f26r	D C	Y	49	
4	f26v	D D	Y	50	
4	f27r	D E	Y	51	
4	f27v	D F	Y	52	
4	f28r	D G	Y	53	
4	f28v	D H	Y	54	
4	f29r	D I	Y	55	
4	f29v	D J	Y	56	
4	f30r	D K	Y	57	
4	f30v	D L	Y	58	
4	f31r	D M	Y	59	
4	f31v	D N	Y	60	
4	f32r	D O	Y	61	
4	f32v	D P	Y	62	
5	f33r	E A	Y	63	
5	f33v	E B	Y	64	
5	f34r	E C	Y	65	
5	f34v	E D	Y	66	
5	f35r	E E	Y	67	
5	f35v	E F	Y	68	
5	f36r	E G	Y	69	

5	f36v	E H	Y	70	
5	f37r	E I	Y	71	
5	f37v	E J	Y	72	
5	f38r	E K	Y	73	
5	f38v	E L	Y	74	
5	f39r	E M	Y	75	
5	f39v	E N	Y	76	
5	f40r	E O	Y	77	
5	f40v	E P	Y	78	
6	f41r	F A	Y	79	
6	f41v	F B	Y	80	
6	f42r	F C	Y	81	
6	f42v	F D	Y	82	
6	f43r	F E	Y	83	
6	f43v	F F	Y	84	
6	f44r	F G	Y	85	
6	f44v	F H	Y	86	
6	f45r	F I	Y	87	
6	f45v	F J	Y	88	
6	f46r	F K	Y	89	
6	f46v	F L	Y	90	
6	f47r	F M	Y	91	
6	f47v	F N	Y	92	
6	f48r	F O	Y	93	
6	f48v	F P	Y	94	
7	f49r	G A	Y	95	
7	f49v	G B	Y	96	
7	f50r	G C	Y	97	
7	f50v	G D	Y	98	
7	f51r	G E	Y	99	
7	f51v	G F	Y	100	
7	f52r	G G	Y	101	
7	f52v	G H	Y	102	
7	f53r	G I	Y	103	
7	f53v	G J	Y	104	
7	f54r	G K	Y	105	

7	f54v	G L	Y	106	
7	f55r	G M	Y	107	
7	f55v	G N	Y	108	
7	f56r	G O	Y	109	
7	f56v	G P	Y	110	
8	f57r	H A	Y	111	
8	f57v	H B	Y	112	
8	f58r	H C	Y	113	
8	f58v	H D	Y	114	
8	f65r	H E	Y	115	
8	f65v	H F	Y	116	
8	f66r	H G	Y	117	
8	f66v	H H	Y	118	
9	f67r	I A	NO		Placeholder for foldout
9	f67r1	I B	Y	119	Also occasionally 120
9	f67r2	I C	Y	121	
9	f67v	I D	NO		Placeholder for foldout
9	f67v2	I E	Y	122	
9	f67v1	I F	Y	123	
9	f68r	I G	NO		Placeholder for foldout
9	f68r1	I H	Y	125	
9	f68r2	I I	Y	126	
9	f68r3	I J	Y	127	
9	f68v	I K	NO		Placeholder for foldout
9	f68v3	I L	Y	128	
9	f68v2	I M	Y	129	
9	f68v1	I N	Y	130	
10	f69r	J A	Y	131	
10	f69v	J B	Y	132	
10	f70r	J C	NO		Placeholder for foldout
10	f70r1	J D	Y	133	
10	f70r2	J E	Y	134	
10	f70v	J F	NO		Placeholder for foldout
10	f70v2	J G	Y	135	
10	f70v1	J H	Y	136	
11	f71r	K A	Y	137	

11	f71v	K B	Y	138	
11	f72r	K C	NO		Placeholder for foldout
11	f72r1	K D	Y	139	
11	f72r2	K E	Y	140	
11	f72r3	K F	Y	141	
11	f72v	K G	NO		Placeholder for foldout
11	f72v3	K H	Y	142	
11	f72v2	K I	Y	143	
11	f72v1	K J	Y	144	
12	f73r	L A	Y	145	
12	f73v	L B	Y	146	
13	f75r	M A	Y	147	
13	f75v	M B	Y	148	
13	f76r	M C	Y	149	
13	f76v	M D	Y	150	
13	f77r	M E	Y	151	
13	f77v	M F	Y	152	
13	f78r	M G	Y	153	
13	f78v	M H	Y	154	
13	f79r	M I	Y	155	
13	f79v	M J	Y	156	
13	f80r	M K	Y	157	
13	f80v	M L	Y	158	
13	f81r	M M	Y	159	
13	f81v	M N	Y	160	
13	f82r	M O	Y	161	
13	f82v	M P	Y	162	
13	f83r	M Q	Y	163	
13	f83v	M R	Y	164	
13	f84r	M S	Y	165	
13	f84v	M T	Y	166	
14	f85r	N A	NO		Placeholder for foldout
14	f85r1	N B	Y	167	
14	f85r2	N C	Y	171	
14	fRos	N D	Y		
14	f85v	N E	NO		Placeholder for foldout

14	f85v2	N F	NO		Transliterated text is part of fRos
14	f85v1	N G	NO		Transliterated text is part of fRos
14	f86r	N H	NO		Placeholder for foldout
14	f86r4	N I	NO		Transliterated text is part of fRos
14	f86r3	N J	NO		Transliterated text is part of fRos
14	f86r6	N K	NO		Transliterated text is part of fRos
14	f86r5	N L	NO		Transliterated text is part of fRos
14	f86v	N M	NO		Placeholder for foldout
14	f86v4	N N	Y	168	
14	f86v6	N O	Y	169	
14	f86v5	N P	Y	172	
14	f86v3	N Q	Y	173	
15	f87r	O A	Y	175	
15	f87v	O B	Y	180	
15	f88r	O C	Y	181	
15	f88v	O D	Y	182	
15	f89r	O E	NO		Placeholder for foldout
15	f89r1	O F	Y	183	
15	f89r2	O G	Y	184	
15	f89v	O H	NO		Placeholder for foldout
15	f89v2	O I	Y	185	
15	f89v1	O J	Y	186	
15	f90r	O K	NO		Placeholder for foldout
15	f90r1	O L	Y	187	
15	f90r2	O M	Y	188	
15	f90v	O N	NO		Placeholder for foldout
15	f90v2	O O	Y	189	
15	f90v1	O P	Y	190	
17	f93r	Q A	Y	191	
17	f93v	Q B	Y	192	
17	f94r	Q C	Y	193	
17	f94v	Q D	Y	194	
17	f95r	Q E	NO		Placeholder for foldout

17	f95r1	Q F	Y	195	
17	f95r2	Q G	Y	196	
17	f95v	Q H	NO		Placeholder for foldout
17	f95v2	Q I	Y	197	
17	f95v1	Q J	Y	198	
17	f96r	Q K	Y	199	
17	f96v	Q L	Y	200	
19	f99r	S A	Y	201	
19	f99v	S B	Y	202	
19	f100r	S C	Y	203	
19	f100v	S D	Y	204	
19	f101r	S E	Y	205	
19	f101r1	S F	NO	205	Transliterated text is included in f101r
19	f101r2	S G	NO	205	Transliterated text is included in f101r
19	f101v	S H	Y		
19	f101v2	S I	NO	206	Transliterated text is included in f101v
19	f101v1	S J	NO	207	Transliterated text is included in f101v
19	f102r	S K	NO		Placeholder for foldout
19	f102r1	S L	Y	208	
19	f102r2	S M	Y	209	
19	f102v	S N	NO		Placeholder for foldout
19	f102v2	S O	Y	210	
19	f102v1	S P	Y	211	
20	f103r	T A	Y	212	
20	f103v	T B	Y	213	
20	f104r	T C	Y	214	
20	f104v	T D	Y	215	
20	f105r	T E	Y	216	
20	f105v	T F	Y	217	
20	f106r	T G	Y	218	
20	f106v	T H	Y	219	
20	f107r	T I	Y	220	
20	f107v	T J	Y	221	
20	f108r	T K	Y	222	

20	f108v	T L	Y	223	
20	f111r	T M	Y	224	
20	f111v	T N	Y	225	
20	f112r	T O	Y	226	
20	f112v	T P	Y	227	
20	f113r	T Q	Y	228	
20	f113v	T R	Y	229	
20	f114r	T S	Y	230	
20	f114v	T T	Y	231	
20	f115r	T U	Y	232	
20	f115v	T V	Y	233	
20	f116r	T W	Y	234	
20	f116v	T X	Y	235	