

**WG2 N5128R**

**L2/20-068R**

***Date: 2020-03-12***

**Title: Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, all Groups (A to Z and AA)**

**Source: Michel Suignard**

**Status: Individual Contribution**

**Distribution: UTC, WG2**

**Replace: N5063 (L2/19-220)**

**Summary:** This document contains a revised draft proposal for the encoding of an extended Egyptian Hieroglyph repertoire based on the taxonomy proposed in document [L2/16-257] and revised from [L2/17-073] with the addition of many more references. This document, as a revision of the original document now, covers all group A to Z and AA. In addition to the existing 1071 hieroglyphs, this proposal contains an additional 6826 code points for a total of 7897 code points. Unlike the previous version (N5603), the document presents the existing Egyptian Hieroglyph block to facilitate the review. For harmonization, the glyphs of the existing block have been modified from the original design and have been part of recent Unicode version code charts; this has introduced some variations that have been questioned by some experts, resulting in some glyph changes for the already encoded block. Therefore, the presentation of the already assigned characters may still evolve.

## Introduction

Egyptian hieroglyphs have evolved from a set slightly above 700 characters during most of the classical period (from Old Kingdom, Middle Kingdom, to New Kingdom) to a much larger repertoire in the Greco-Roman period. In that era, the number of hieroglyphic characters has expanded to a number above 7000. Typically, that larger set is known as Ptolemaic, but this is a simplistic assumption as many of these extensions have been found to be in use in the classical period as well.

The explosion in glyph diversity was also accompanied by an increase in the number of phonetic values (phonemes) that could be assigned to a single glyph, up to twenty or even more readings. Overall, deciphering late usage of the script has proven to be challenging.

This document does not pretend to describe in detail the various complexities of the extended Egyptian Hieroglyphs and its usage. There are many publications on that subject, the bibliographic references section has been extended in the recent versions of this document to covers some related documents on that subject.

A previous document: [L2/16-257] established the base for the encoding model. The sign taxonomy is based on the [IFAO] classification (Institut Français d'Archéologie Orientale du Caire), roughly based on the same large groups as the Gardiner classification but with a secondary level allowing finer and more manageable grouping. Thus, a database has been created with all Hieroglyphica glyphs/characters, along with added hieroglyphs through the study of references.

While some of the proposed elements could be considered 'quadrat' structures, the emphasis of the proposal is to encode 'atomic' characters, i.e. characters that cannot be easily rendered as discrete elements because it involves ink collision and complex reshaping of the components.

Depending on parallel discussion concerning documents such as [L2/16-210], and [L2/17-112R], some characters may be removed from this proposal, if consensus is reached to use an alternate mode to encode some of the proposed characters.

All Egyptian hieroglyphs already encoded in the existing Egyptian Hieroglyphs block (U+13000..U+1342F) have been entered into the database using the new taxonomy attributes.

The intent of this proposal is to set the base for the encoding of the extended set, allowing Egyptologists to communicate data in a unified encoding platform. Then each element of the extended set can be indexed in a database where additional information can be conveyed, related to the phoneme, semantics, sources, description, etc. In many aspects, this is like the work done for the Han ideographs where many ancillary data elements are associated with each Han characters (commonly known as the 'Unihan' database).

The author is also aware that there is a risk of encoding quasi duplicates (i.e. characters that may not carry any significant distinction from another proposed) or an already encoded character, or even ghost characters (i.e. characters for which attestation is in doubt at present or could be found not attested in the future due to research error). However, a meticulous exploration of the subject could take decades, so it seems reasonable to encode some potential superfluous characters. Through the usage of a database, researchers can always qualify set elements (that is, define legitimate characters). This is quite like the challenges and solutions used in the context of CJK Ideographs which is a growing set of more than 100,000 code points with many of the same challenges.

It should be noted that the sign taxonomy used in this document is only meant to improve the visual classification of these signs to facilitate the search and update of the content. It is not meant to be a taxonomy covering all functions that can be fulfilled by the signs. This is clearly beyond the scope of this document. A good study of such taxonomy is done in 'The Hieroglyphic Sign Functions' [Taxonomy]. Some of the classification terminology from that document is introduced later here.

## Sign Taxonomy

The proposed sign taxonomy uses a three-level classification. The higher level is a combination of the Gardiner A-Z (and Aa) classification and the IFAO chapter classification (I to XXX in roman notation). While IFAO has a few more items, they can be easily mapped into existing Gardiner groups (for example the Gods (Chapter III) and Goddesses (Chapter IV) can be combined in the Gardiner group C (Anthropomorphic Deities). The following is the list of the first level groups and their relationship with the IFAO groups:

<b>Gardiner groups</b>	<b>IFAO (translated from French)</b>
A. Man and his occupations	I. Men and monarchs
B. Woman and her occupations	II. Women and monarchs
C. Anthropomorphic deities	III. Gods IV. Goddesses
D. Parts of the human body	V. Human body parts
E. Mammals	VI. Mammals
F. Parts of mammals	VII. Mammal body parts
G. Birds	VIII. Birds
H. Parts of birds	IX. Bird parts
I. Amphibious animals, reptiles, etc.	X. Reptiles, amphibians
K. Fishes and parts of fishes	XI. Fishes and parts of fishes
L. Invertebrate and lesser animals	XII. Insects and arachnids
M. Trees and plants	XIII. Plants
N. Sky, earth, water	XIV. Sky, earth, water
O. Buildings, parts of buildings, etc.	XV. Edifices and parts of edifice
P. Ships and part of ships	XVI. Boats and parts of boat
Q. Domestic and funerary furniture	XVII. Everyday and funeral furniture
R. Temple furniture and sacred emblems	XVIII. Temple furniture
S. Crowns, dresses, staves, etc.	XIX. Crowns

Gardiner groups	IFAO (translated from French)
	XX. Jewels, clothes, staves
T. Warfare, hunting, butchery	XXII. Warfare, hunting, fishery, butchery
U. Agriculture, crafts, and professions	XXI. Agriculture and shop tools
V. Rope, fiber, baskets, bags, etc.	XXIII. Rope, baskets, bags
W. Vessels of stone and earthenware	XXIV. Vases
X. Loaves and cakes	XXV. Bread loaves
Y. Writings, games, music	XXVI. Writings, games, music
Z. Strokes, signs derived from Hieratic, geometrical figures	XXVII. Geometric shapes
AA. Unclassified	XXVIII. Ill-defined signs

Notes:

- The order of the A-Z and I to XXVIII lists is identical except for the 2 groups T-U and XXI-XXII.
- IFAO Chapter XXIX (Uncertain identity signs) and Chapter XXX (Conventional signs) are not used in the taxonomy because they have no known references in the Hieroglyphica set and are seldom used by other references.
- Many characters originally from the group 'AA.XXVIII Unclassified Ill-defined signs' have been moved to other groups when their identity could be clarified. Some members originally from the IFAO group XXIX have also been reclassified.

To facilitate the transition from the current Gardiner sign taxonomy, it seemed preferable to keep the same naming convention A to Z for the higher level, to eschew J, and to put the uncertain code points in the group 'AA'.

The second level uses the IFAO classification within each IFAO chapter to further enumerate the content of each of these groups. For example, the IFAO classification for 'A: Man and his Occupations' has the following categories (translated from French):

Men and Rulers (homme et souverains)










1. Man seated or kneeling empty handed (homme assis ou agenouillé les mains vides)
2. Man standing empty handed (homme debout les mains vides)
3. Man, head down (homme, la tête en bas)
4. Man adoring or bent (homme adorant ou penché)
5. Man on the ground or in water (homme à terre ou dans l'eau)
6. Man standing, holding a staff (homme debout, tenant un bâton)
7. Man seated or kneeling, holding something (homme assis ou agenouillé, portant quelque chose)
8. Man standing, holding something (homme debout, portant quelque chose)
9. Man seated, pouring water (homme assis, versant de l'eau)
10. Man standing, pouring water (homme debout, versant de l'eau)
11. Man hiding (homme se cachant)
12. Man working (homme au travail)
13. Shepherd and porter (berger et portier)
14. Man carrying a bundle (homme portant un baluchon)
15. Man standing, composed with a hieroglyphic sign (homme debout, en composition avec un signe hiéroglyphique)
16. Man or god holding a weapon (homme tenant une arme)
17. Soldier (soldat)
18. Prisoner and enemy (prisonnier et ennemi)
19. Dancer and acrobat (danseur et acrobates)
20. Musician (musicien)
21. Man and animal (homme et animal)
22. Man in a boat (homme dans un bateau)

23. Child (enfant)
24. Dwarf (nain)
25. Noble seated on a chair (notable assis sur un siège)
26. King or god seated without crown (Roi ou dieu assis sans couronne)
27. King or god standing without crown (Roi ou dieu debout sans couronne)
28. King or god wearing the white crown (Roi portant la couronne blanche)
29. King or god wearing the red crown (Roi portant la couronne rouge)
30. King or god wearing another crown (Roi portant une autre couronne)
31. Dead person kneeling (défunt agenouillé)
32. Mummy (momie)
33. Statue (statue)

This categorization is used in the new taxonomy, with just one additional value (34. Man, varia) added to cover signs that are not attributable to any of the previous sub-groups. Once these members are created, they are immutable, i.e. they cannot be removed or renumbered. However, new members can be created in each of the A to Z and AA first level groups if such a need arises. Finally, the glyphs from Hieroglyphica have been mapped into each of these 2<sup>nd</sup> level sub-groups to create the 3<sup>rd</sup> level content. The format use the following syntax: ((A-IK-Z){1}|AA)-(0-9){2}-(0-9){3}. Examples are A-01-001, M-21-024, AA-08-006.

The following table shows an example for the sub-group A11 made of 9 characters: (1<sup>st</sup> column shows the new index, 2<sup>nd</sup>: representative glyph; 3<sup>rd</sup>: UCS code point (existing or draft); 4<sup>th</sup>: IFAO index. The glyphs which are shown as black silhouettes are not currently encoded. They were created using a graphic snapshot of the IFAO documentation.

**Table for A11: Man Hiding (only showing IFAO source)**

index	glyph	UCS	IFAO
A-11-001			15;2
A-11-002		13005	15;3
A-11-003		13548	15;4
A-11-004			15;5
A-11-005			15;6
A-11-006			15;7
A-11-007		13004	
A-11-008		13549	
A-11-009		1354A	

## Sign taxonomy applied to the Hieroglyphica set

Having defined a sign taxonomy, the next step was to validate it by applying it to the target set which is the Hieroglyphica set, augmented by extensions found in implemented fonts (the set is the base of document L2/16-079), and extensions made to cover other sources.










The following index and sources were analyzed:

- Gardiner, 1977
- Hieroglyphica index (including addition made by Jochen Hallof in reporting Dendara attestation)
- Hieroglyphica extension (various sources)
- Douros extensions (Aegyptus)
- JSesh word processor, version 7.3.2

- Unikemet (as presented in existing Unicode Egyptian Hieroglyphs block)
- Dendara, Le fonds hiéroglyphique au temps de Cléopâtre, Sylvie Cauville, 2001
- Catalogue de la fonte hiéroglyphique de l'imprimerie IFAO, 1983
- Christian Leitz – Einführungen und Quellentexte zur Ägyptologie
- Kurth Einführung ins Ptolemaische, 2009
- Hornung & Schenkel (2007, last modified in 2015)
- Möller extensions (as conveyed in L2/16-250)
- Richmond extensions (as conveyed in L2/16-303), see <http://www.egpz.org/docs/AnalysisOfSomeMdCCodedTexts.pdf>
- Jochen Hallof attestation, referencing: Dendara, Les chapelles Osiriennes, Sylvie Cauville, 1997-2012
- Valeurs Phonétiques, 1988

Based on source attestation a weight index was created to denote the number of sources found. In some rare cases, where multiple sources described a character not found in an existing font, the glyph was created as an extension to the Hieroglyphica set. At this point all entries with a weight entry value of '1' or more may be proposed for encoding (if not already encoded in the existing block).

**Table for A11: Man Hiding (showing all sources with at least a member for this group)**

index	glyph	UCS	HG	JSesh	Gard	Unik.	Cauv.	IFAO	Leitz	Kurth	H&S	JH	VP	W
A-11-001								15;2					A;923	2
A-11-002		13005	A5A	A5A	A5*	A005A	9;4	15;3	153;4	1;35	A;1851	D70	A;925	9
A-11-003		13548	A5C	A5C				15;4				D16	A;927	3
A-11-004								15;5					A;931	2
A-11-005								15;6					A;933	2
A-11-006								15;7					A;929	2
A-11-007		13004	A5	A5	A5	A005			153;4	1;35	A;1850	D1		6
A-11-008		13549	A5B	A5B							A;1852			1
A-11-009		1354A	A421	A421									A;921	1

Legend:

- Index: category value,
- Glyph: representative glyph, if it is a black silhouette it has not been yet encoded.
- UCS: proposed or existing code points,
- HG: Hieroglyphica index,
- JSesh: JSesh index,
- Gard: Gardiner 1957 index,
- Unikemet index,
- Cauv.: Dendara by Sylvie Cauville, index,
- IFAO: index,
- Leitz: index,
- Kurt: index,
- H&S: Hornung & Schenkel index,
- JH: Jochen Hallof reference index,
- VP: Valeurs Phonétiques index,

- W: Weight value

Note that, because IFAO was the base for the taxonomy, many indexes only referenced by IFAO or VP were kept as reserved entries. For example, the value A-11-001, A-11-004, A-11-005, and A11-11-006 corresponding to IFAO entries: 15;2, 15;5, 15;6, and 15;7 are reserved. However, no code points are allocated for those reserved entries.

Note also that 3 sources are not present in the table above because they are not used for these 9 characters:

- Möller extensions,
- Hieroglyphica index extension,
- Richmond extensions.

Hieroglyphica is by far the biggest contributor to the repertoire but shares approximately 6800 signs with JSesh; this represents the largest intersection between two sources of this repertoire. The union of these two sets has become a de facto reference for identifying hieroglyphic signs in other publications when there is a desire to expand the sign repertoire beyond the basic Gardiner set. Some examples are 'Le fonds hiéroglyphique au temps de Cléopâtre' Sylvie Cauville, Cybèle, [Cauville], 'le Dictionnaire des Hiéroglyphes', Yvonne Bonnamy, Actes Sud [Dictionnaire], but there are many more. In addition, since Hieroglyphica was first established, the catalogue has expanded significantly based on more recent work by Jochen Hallof in identifying signs used in various Dendara based publications.



A common criticism of repertoires like Hieroglyphica is that they represent a modern abstraction of what is inherently a much larger repertoire of original classic attestation as found in stones, wood carving, etc. Every publication has struggled with this issue by creating their own abstract reference to represent as a single sign for what are different instances of slightly different carved signs. A good example is the sign N37A. In the Gardiner 1953 Supplement, it is typically represented by the following glyph:



In Hieroglyphica, the same sign is represented by the following glyph:



But in looking at Dendara pictorial evidences, it is possible to observe that the abstraction represents both forms. But, looking at Valeurs Phonétiques page 469, the two glyphs are given different values:

623		<i>mr</i> <i>n</i> <i>ntr</i> <i>h</i> <i>hr</i> <i>h</i> <i>š</i>	Loret, Dümichen, Devéria, Junker, Junker, Sauneron, Lepsius,	<i>Manuel, n° 502.</i> <i>ZAS 3, 1865, p. 58.</i> <i>BE V, 1897, p. 71.</i> <i>Sprachliche, p. 15.</i> <i>Pylon, p. 52,6, n.2.</i> <i>Esna VIII, p. 164.</i> <i>Königsbuch, p. 175.</i>	
627	2007 	<i>mr</i> <i>n</i> <i>nt</i> <i>h</i> <i>š</i> <i>š</i> <i>šw</i>	Fairman, Drioton, Drioton, Fairman, Fairman, Alliot, Daumas,	<i>BIFAO 43, 1945, p. 71, n1.</i> <i>A.S.A.E. 44, 1944, 119 C.</i> <i>A.S.A.E. 44, 1944, p. 119.</i> <i>A.S.A.E. 43, 1943, p. 238.</i> <i>BIFAO 43, 1945, p. 77.</i> <i>Culte I, p. 248.</i> <i>Mammisis, p. 305, n.4.</i>	278,8

At some point, the author was tempted to unify these two glyphs, using the Hieroglyphica glyph because it seems that the Dendara attestations favor the original Hieroglyphica glyph. But because the sets of graphemes attached to the two signs are not identical, it seems prudent to create two code points, even if in most cases they are probably not distinguishable.

# Encoding Proposal

The new sign taxonomy makes it much easier to insert new code points in a meaningful staged approach. While the encoding of similar characters may not be contiguous (that is already not possible because 1071 of them were encoded in the current Egyptian Hieroglyphs block), the taxonomy will make easy to identify later updates as part of the 'same' sub-groups.

The current set was created using the methodology explained above and is proposed as follows:

- 1342F: 1 code point to complete the existing Egyptian Hieroglyphs block (U+13000..U+1342F): EGYPTIAN HIEROGLYPH AA030A (new taxonomy index: AA-18-006)
- 13440..143ED: 4014 code points in a new block Egyptian Hieroglyphs Extended-A (U+13440..U+143FF) covering the following sections:
  - A.I. Men and monarchs
  - B.II Women and monarchs
  - C.III Gods
  - C.IV Goddesses
  - D.V Human body parts
  - E.VI Mammals
  - F.VII Mammal body parts
  - G.VIII Birds
  - H.IX Bird parts
  - I.X Reptiles, amphibians
  - K.XI Fishes and part of fishes
  - L.XII Insects and arachnids
  - M.XIII Plants, including the subgroups from M01. Tree to M08. Lotus flower
- 14680..1517A: 2811 code points in a new block Egyptian Hieroglyphs Extended-B (U+14680..U+151FF) covering the remaining sections:
  - M.XIII Plants, including the subgroups from M09. Composite bunch to M29. Vegetal elements, varia
  - N.XIV Sky, earth, water
  - O.XV Edifices and parts of edifices
  - P.XVI Boats and parts of boat
  - Q.XVII Domestic and funerary furniture
  - R.XVIII Temple furniture
  - S.XIX Crowns
  - S.XX Jewels, clothes, staves
  - T.XXII Warfare, hunting, fishery, butchery (out of order before XXI)
  - U.XXI Agriculture and shop tools
  - V.XXIII Rope, baskets, bags
  - W.XXIV Vessels of stone and earthenware
  - X.XXV Loaves and cakes
  - Y.XXVI Writings, games, music
  - Z.XXVII Geometric signs, hieratic signs, strokes
  - AA.XXVIII Unclassified signs

These 6826 code points augment the 1071 characters that are already encoded in the existing Egyptian Hieroglyphs block in the range U+13000-U+1321F.

Characters for which there are no representative font glyphs have no reserved code points. For example, considering the set A11 previously described, the current code chart is defined as follows:

## A11. Man hiding

A-11-002 referenced as A005A

→ 13005  egyptian hieroglyph a005a

13548  EGYPTIAN HIEROGLYPH A-11-003

A-11-007 referenced as A005

→ 13004  egyptian hieroglyph a005

13549  EGYPTIAN HIEROGLYPH A-11-008

1354A  EGYPTIAN HIEROGLYPH A-11-009

The code chart is not reserving code points for A-11-001, A-11-004, A-11-005, and A-11-006. This could be easily changed if preferred. There are currently over 2000 characters which fall in that category. Reviewers may look at the database extract provided as an appendix to this document to possibly augment the currently proposed set by proposing additional characters (currently shown as black silhouette pictographs). Reviewers can request that specific silhouetted characters be added.

A full code chart for the existing block (Egyptian Hieroglyphs) and the newly proposed 6826 characters in two new blocks (Egyptian Hieroglyphs Extended-A and Egyptian Hieroglyphs Extended-B) is appended to this document.

## Unicode properties

### UnicodeData.txt:

```
1342F; EGYPTIAN HIEROGLYPH AA030A;Lo;0;L;;;;N;;;;
13440;EGYPTIAN HIEROGLYPH A-01-003;Lo;0;L;;;;N;;;;
...
143ED; EGYPTIAN HIEROGLYPH M-08-043;Lo;0;L;;;;N;;;;
14680; EGYPTIAN HIEROGLYPH M-09-001;Lo;0;L;;;;N;;;;
...
1517A;EGYPTIAN HIEROGLYPH AA-22-020;Lo;0;L;;;;N;;;;
```

**Script value:** Egyptian\_Hieroglyphs

All other property values are similarly derived from values assigned to existing Egyptian Hieroglyphs and therefore the information is not duplicated here.

Ideally, sorting order should be done according to the new sign taxonomy order, but it is not a hard requirement.

## Functional Taxonomy

The taxonomy used in this document only aims to facilitate the search of specific signs in this large set and is mostly based on their visual appearance. It is also designed to facilitate additions to the set. In contrast, a functional taxonomy would take into consideration the functions of these signs. Originally, the functional taxonomy used by many Egyptologist to classify the hieroglyphic signs was using 3 base classes: Phonogram, Ideogram and Determinative which reflected their main use. One of the problems with that classification is that Ideograms have also an inherent phonetic property making somewhat difficult to separate Ideograms and Phonograms. The [Taxonomy] document suggests the following groups:

	+ SEMOGRAM		- SEMOGRAM
AUTONOMOUS	Pictogram	Logogram	Phonogram
NON-AUTONOMOUS	Classifier	Radicogram	Interpretant
	- PHONEMOGRAM		+ PHONEMOGRAM



It classifies the signs according to 3 dimensions (content: SEMOGRAM, linguistic form: PHONEMOGRAM, and autonomy: AUTONOMOUS or NON-AUTONOMOUS). It further refines the usual classification of logogram, ideogram, and determinative. The following description is derived from the [Taxonomy] document:

- Pictogram: not linked to a fixed sequence of phonemes,
- Logogram: grapheme that refers to entire words,
- Phonogram: phoneme or combination of phonemes,
- Classifier: morpheme that occurs at the end of word and gives some indications about the semantic classification of a lexeme,
- Radicogram: grapheme that points at the same time to some form and some content but is not able alone to refer to an autonomous lexeme,
- Interpretant: non-autonomous grapheme that interprets the phonemic value of other graphemes.

It is expected that eventually these classes will be used to describe and study all signs; many of those signs can have many classes depending on context. For example, the content of 'Valeurs Phonétiques' should be evaluated according to all these classes.

## Database

One of the essential tools of this Egyptian hieroglyphs inventory is a database containing all listed entries, encoded or not, collecting all ancillary data concerning each of these entries. It is a work in progress as new elements are added for each of these entries. Currently, these elements concern source references but could be extended to cover additional information. The reference [Database] is a PDF representation of that database. It is available as a separate file associated to this document.










The current format of the database is an Excel spreadsheet with the following columns:

- Taxonomy index, only used as a guide to the following rows. When present, no other columns are defined. For other rows, the column contains a provisional PUA code to be used in a future version of the base font for this work.
- Catalog number, using the taxonomy format described above;
- Representative glyph, it may be a font glyph or a picture;
- Code point (dbCode), any value below U+1342F is an actual Unicode code point (U+1342F is newly proposed), values above 13430 are internal to the database and are only used by the font displaying the glyph in the previous column. The intent is to eventually move these values to either the proposed encoding value or PUA values in the U+F0000-U+FFFFFF range. These are not UCS codes.
- Hieroglyphica index, based on Gardiner naming scheme. These values are generally found in all derived work. The number of these indexes have augmented with successive publications of the Dendara evidences.
- Hieroglyphica extended index, these are not part of the original Hieroglyphica, and can be found in Hieroglyphica derived fonts. The author has also added values in this area to represent glyphs not available before but found in some sources. Some provisional values (noted as DExxx) used in Dendara referencing are also recent additions.
- JSesh sign index, based on the JSesh word processor version 7.3.2.
- George Douros Aegyptus font index extension; these are mostly used to represent hieroglyphs that were not in the original Gardiner set, or in case where Hieroglyphica glyphs were different from Gardiner for the same reference.
- JSesh sign index; JSesh is a word processor for ancient Egyptian hieroglyphic texts. Its sign collection uses a syntax which matches most of the Hieroglyphic index.

- Gardiner index based on publications from 1928 to 1953. Some indexes were collected from the 1957 Egyptian Grammar publication. All these references are already encoded in the original Egyptian Hieroglyphs block.
- Unikemet database index, based on document [L2/06-355] (same as ISO/IEC SC2/WG2 N3182) and [L2/07-097] (same as WG2 N3237).
- Dendara, Sylvie Cauville, 2001 [Cauville]. This reference uses both Hieroglyphica glyph id and IFAO index, which makes it very useful to confirm these mappings. The book references are in “page number; item number in page”.
- Catalogue de la fonte Hiéroglyphique de l’imprimerie de l’I.F.A.O., 1983 [IFAO]. This column represents IFAO entries using page number and item number within the page. Given the different styles used between Hieroglyphica glyphs and symbols used in IFAO, matching the symbols has been challenging.
- Einführungen und Quellentexte zur Ägyptologie, Louise Gestermann und Christian Leitz [Leitz]. This reference uses Hieroglyphica glyphs and Id. The book references are in “page number; item number in page”.
- Einführung ins Ptolemäische, Dieter Kurth [Kurth]. This reference uses glyphs very similar to Hieroglyphica but with additions. Some of these additions were added to the repertoire. The book references are in “page number; item number in page”.
- Hornung & Schenkel (2007, last modified in 2015). The references use the 4-digit numbers found in each A-Z and AA categories in the source in the form ‘A;0100’. Sometimes, only glyphs were provided without a clear numeric reference, in those case incomplete references using ‘X’ to denote approximation. For example, the reference ‘O;06XX’ indicates an entry associated with references found in the range O;0600 to O;0699, but not having an explicit numeric value.
- Möller extensions (as conveyed in L2/16-250) <http://www.unicode.org/L2/L2016/16250-n4741-moller-egyptian.pdf>, references that could be identified in Hieroglyphica were added as referenced sources in the database. The references used for that column are the ones (slightly edited for concision) provided in annotation of the name list in L2/16-250. For example, that document proposes EGYPTIAN HIEROGLYPH A084 with annotation ‘Moeller a020bis v3’, the reference column in the database has the value ‘a20bis v3’.
- Bob Richmond extensions (as conveyed in [L2/16-303]), see <http://www.egpz.org/docs/AnalysisOfSomeMdCCodedTexts.pdf>, the references used are the ‘CODE’ values in Table 1.
- Jochen Hallof, referencing: Dendara, Les chapelles Osiriennes, Sylvie Cauville, Volumes X to XV, 1997-2012 and La Porte d’Isis, 1999. The index uses the values Dn (n is the number showing the number of attestations in the documents; for example D1 indicates an hapax). The value ‘x’ indicates a Hieroglyphica glyph for which another attestation was not yet found but is likely to be eventually found. The author has found that in the progression of this work the amount of this yet unattested set has shrunk significantly, justifying encoding these glyphs.
- Valeur Phonétiques des signes Hiéroglyphiques d’Époque Gréco-Romaine (VP), Montpellier, 1988. The index uses the categories A-Z. For the current subsets, because God and Goddess use separate the same index space in VP, the goddess value has been shifted by 1000 to create unique values for the goddess entries (C;1 -> C;1001).
- The actual value associated with a VP index, multiple values are separated by ‘;’. These values are Latin transliteration of the graphemes of these elements without determination of their functional taxonomy classification. This value does not count for the weight index calculation.
- Weigh index, this value is computed by accumulating the number of references found for a given entry. Only references belonging to Gardiner, and columns on the left are counted.
- Comment column, presenting remarks on the reference, note that the ‘~~~~’ notation indicates some imperfect matching.

The following table shows the ‘valeurs phonétiques’ values for the group A11 (other values are shown in the previous table):

**Table for A11: Man Hiding (showing Valeurs Phonétique values)**

index	glyph	UCS	VP values	Note
A-11-001			<i>h3p</i>	Not proposed
A-11-002		13005	<i>imn;rm</i>	Already encoded
A-11-003		13547		
A-11-004			<i>imn;nm</i>	Not proposed
A-11-005			<i>thn</i>	Not proposed
A-11-006			<i>imn;thn</i>	Not proposed
A-11-007		13004		Already encoded
A-11-008		13539		
A-11-009		1353A		

Other columns are planned to be added to the database, such as the following:

- Description – free form description of the glyph
- Components – list of all glyph components
- Classes of use according to the functional taxonomy
- Variants – similar entries

## Current issues

1. How to maintain IFAO taxonomy principle for extension? IFAO provides some principle for compound signs to determine which part of the compound sign should be used to classify the glyph:
  - Container preempts sign contained inside,
  - Signs are classed in order of reading order (difficult to maintain by this author),
  - Vertical sign preempts horizontal sign

Typically, the process has been to look at existing cases to determine an underlying logic, not always explicitly described in IFAO. Even in few cases, the IFAO classification has not been followed because it was unexpected.

2. Encoding of the IFAO entries not yet represented in mode fonts. It would be tempting to add glyphs for IFAO entries which have additional attestation such as in Valeurs Phonétiques, especially when these entries have distinct graphemes.
3. Consistency of the classification. There are some arbitrary decisions to be made in categorizing characters. For example, there is not always a consensus in determining whether a character represents an animal, or an abstraction such as a pottery with animal design. Especially for hapax, the categorization can be arbitrary. At this stage of the proposal, re-categorization based on additional evidence is perfectly acceptable.
4. What to do with the AA categories (Unclassified signs)? As research is progressing, more and more of these characters are re-classified into other ‘classified’ categories. It would seem natural to move these characters into their latest classification. This has been done for many of them and could use some feedback from experts.

## Correlation with other proposals to extend Egyptian hieroglyphs

Along with this work, other documents have been created, proposing smaller additions to existing repertoire. For example, [L2/16-250], authored by Michael Everson, is proposing an extension to the Egyptian Hieroglyphs by adding 146 characters to the Egyptian Hieroglyphs (one code point in the existing block, the other 145 in a new extended block). The extended set is made up of 125 characters included in the usual Gardiner index (A-Z, AA) and 21 'uncertain' characters. A vast majority of the first 125 characters are already included in both the Hieroglyphica set and IFAO, and therefore part of this encoding proposal.

The same document is also advocating for the continuation of the same naming scheme used in the original Egyptian Hieroglyphs block. The IFAO catalog in its preamble shows the limitation of such a naming scheme when the set goes from 600-1000 to 7000 or 8000. Some of that rationale is repeated here in the introduction of the new sign taxonomy, and therefore that naming scheme is not used in this proposal.

In a similar fashion, a document created by Bob Richmond, see [L6/16-307] proposed to add 200 code points, based on MdC and Hieroglyphica, and again using a naming scheme based on the original Gardiner list. These 200 characters have been referenced as sources in the current proposal, using their 'CODE' column value as reference value. In some cases, such as mirrored images, new glyphs were created.

## Conclusion

The work presented in this document is still preliminary and some of the information is presented in a rough format. However, the base is now a spreadsheet containing all the pertinent information which can be used to extract correlation points and can be augmented easily with ancillary data to make encoding decision easier as well as facilitate search in this vast set.

## References

### Repertoire:

[L2/06-355]: Source for the encoding of Egyptian Hieroglyphs (WG2 N3182), Michael Everson, 2006-10-29, <http://www.unicode.org/L2/L2006/06355-n3182-egyptian-sources.pdf>

[L2/07-097]: Proposal to encode Egyptian Hieroglyphs in the SMP of the UCS, Michael Everson and Bob Richmond, 2007-04-10 (the content was slightly modified for incorporation in Unicode) <http://www.unicode.org/L2/L2007/07097-n3237-egyptian.pdf>

[L2/16-079], Preliminary draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2016-04-11, <http://www.unicode.org/L2/L2016/16079-hieroglyphs.pdf>

[L2/16-257], Source analysis of an extended Egyptian Hieroglyphs repertoire (Hieroglyphica based), Michel Suignard, 2016-09-19, <http://www.unicode.org/L2/L2016/16257-n4751-hieroglyphs-new.pdf>

[L2/16-250], Preliminary proposal to encode Möller's Egyptian Hieroglyphs in the SMP (WG2 N4741), Michael Everson, 2016-09-14, <http://www.unicode.org/L2/L2016/16250-n4741-moller-egyptian.pdf>

[L2/16-307], Towards an Expansion of the Unicode Hieroglyph repertoire, Bob Richmond, 2016-10-28 <http://www.unicode.org/L2/L2016/16307-hieroglyph-repertoire.pdf>, and linked document, Analysis of Unicode Egyptian hieroglyphs in a collection of Mdc-coded transcriptions: <https://github.com/HieroglyphsEverywhere/Docs/blob/master/Archive/AnalysisOfSomeMdCCodedTexts.pdf>.

[L2/17-073], also WG2 N4788, New draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2017-03-19, <http://www.unicode.org/L2/L2017/17073-n4788-hieroglyphs.pdf>

[L2/17-415], also WG2 N4924, Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2017-11-27, <https://www.unicode.org/L2/L2017/17415-n4924-hieroglyphs.pdf>

[L2/18-165], also WG2 N4944, Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, Michel Suignard, 2018-04-17, <http://www.unicode.org/L2/L2018/18165-n4944-hieroglyphs.pdf>

[L2/19-165], also WG2 N5063, Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, / First Tranche: Human, God and Goddess / Second Tranche: Human parts, Mammals, Mammal parts, Michel Suignard, 2019-06-20, <http://www.unicode.org/L2/L2018/18165-n4944-hieroglyphs.pdf>

## Structure

[L2/16-210R]: A system of control characters for Ancient Egyptian hieroglyphic text (updated version), Mark-Jan Nederhof et al, 201-01-25 <http://www.unicode.org/L2/L2016/16210r-egyptian-control.pdf>

[L2/16-232]: Preliminary analysis of Egyptian Hieroglyph quadrat type, Andrew Glass, 2016-08-04, <http://www.unicode.org/L2/L2016/16232-quadrat-types.pdf>

[L2/17-112R]: A method for encoding Egyptian quadrats in Unicode, Andrew Glass, et al, 2017-05-12, <http://www.unicode.org/L2/L2017/17112r-quadrat-encoding.pdf>

## Other sources (beyond those mentioned in repertoire references above)

[Leitz] Einführungen und Quellentexte zur Ägyptologie, Louise Gestermann und Christian Leitz, Band 2, 2004, LIT, ISBN 3-8253 7340-4

[Kurth] Einführung ins Ptolemäische, Eine Grammatik mit Zeichenliste und Übungsstücken, von Dieter Kurth, Teil 1, Backe-Verlag, 2009

Hornung & Schenkel (2007, last modified in 2015), Zeichenliste, Prof. Dr. E. Hornung, edited by Stéphane Polis.

[Hieroglyphica] Hieroglyphica – Sign List, Nicholas Grimal, Jochen Hallof, Dirk van der Plas, 2<sup>nd</sup> edition, 2000

[Dictionnaire] Dictionnaire des Hieroglyphes, Yvonne Bonnamy, Troisième Edition. Actes Sud. ISBN 978-2-330-12518-9

[JSesh] Rosmorduc, Serge. (2014). *JSesh Documentation*. [online] Available at: <http://jseshdoc.qenherkhopeshef.org> [Accessed Feb 7<sup>th</sup> 2020]

[IFAO] IFAO, Catalogue de la fonte hiéroglyphique de l'imprimerie de l'I.F.A.O., Institut Français d'Archéologie Orientale du Caire, 1983, IF607, SEVPO, Paris, France

[Cauville] Dendara, Le fonds hiéroglyphique au temps de Cléopâtre, Sylvie Cauville, 2001, Cybèle, ISBN 2-9516758-0-1

[Dendara X to XV] Le Temple de Dendara, Les chapelles Osiriennes, Sylvie Cauville, Institut Français d'Archéologie Orientale, Tomes X to XV, Le Caire, 1997-2012 (composition hiéroglyphique Jochen HALLOF et Hans VAN DEN BERG). X: ISBN 2-7247-0199-2, XI: ISBN 2-7247-0279-4, XII: ISBN 978-2-7247-0460-0.

[Dendara PI] Le Temple de Dendara, La Porte d'Isis, Sylvie Cauville, Institut Français d'Archéologie Orientale. Le Caire, 1999 (composition hiéroglyphique Jochen HALLOF et Hans VAN DEN BERG)

[ValPhon] Valeurs Phonétiques des signes hiéroglyphiques d'époque Gréco-Romaine, Préface de Jean-Claude Goyon, 1988, Université de Montpellier.

[Taxonomy] The Hieroglyphic Sign Functions, Suggestions for a Revised Taxonomy, Stéphane Polis, Serge Rosmorduc, Open Repository and Bibliography (ORBI), University of Liège, link:  
[https://orbi.uliege.be/bitstream/2268/186322/1/Polis\\_Rosmorduc\\_2015\\_FSloprieno\\_Bd.I-1.pdf](https://orbi.uliege.be/bitstream/2268/186322/1/Polis_Rosmorduc_2015_FSloprieno_Bd.I-1.pdf)

## Acknowledgements

The author wants to thank the following persons for their contribution to the effort: Debbie Anderson and Ken Whistler for their kind feedback to the previous drafts, Jochen Hallof and Barbara Richter for their contribution to the determination of the mapping to the existing encoded Egyptian hieroglyphs as well as the additional extensions beyond the original Hieroglyphica set. In addition, Barbara Richter and Debbie Anderson provided many documents helping the creation of this draft proposal. Important other resources were the original Egyptian Hieroglyphs proposal created by Michael Everson, new material created by Bob Richmond, as well as the Aegyptus font created by George Douros.

The proposal also benefited from discussions and ideas exchanged with many other Egyptologists during meetings held in Cambridge, UK in July 2016, London, UK in June 2018, Tournan, France in June 2019.

This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.