

Archive of early EGPZ-related material

These notes were originally given in web pages from www.gameset.com. Some minor editorial changes have been made to web links etc. This material is now superseded by information on www.egpz.com and is provided for background interest only.

Bob Richmond (May 2007).

Standards for Computer Software and Egyptology (February 2005)

The closest to a standard for coding Ancient Egyptian as information suitable for computer processing remains *Manuel de codage des textes hiéroglyphiques en vue de leur saisie sur ordinateur* (Jan Buurman, Nicolas Grimal, Jochen Hallof, Michael Hainsworth and Dirk van der Plas, Informatique et Egyptologie 2, Paris 1988). This is usually known just as Manuel de Codage or simply MdC. All the current egyptological software follows this 'standard' to some degree or other, sometimes with extensions. Surprisingly, the 'standard' itself does not appear to be published on the Internet since the original Centre for Computer-Aided Egyptological Research, Utrecht (CCER) website lapsed. (what claims to be a copy of a CCER revised version can be found [here](http://perso.wanadoo.fr/thotweb/encyclopedie/codage.htm) [http://perso.wanadoo.fr/thotweb/encyclopedie/codage.htm]. The new CCER website is [here](http://www.ccer.nl/) [http://www.ccer.nl/]).

The most comprehensive assignment of codes to individual signs is given in *Hieroglyphica* (Nicolas Grimal, Jochen Hallof, Dirk van der Plas, Utrecht/Paris, 1993). This does not exactly follow the Gardiner/Oxford coding on which it is based but does succeed in associating Gardiner style codes with thousands of good quality consistent drawings. It also departs significantly from Gardiner by intermixing a large number of signs from the Ptolemaic period with Middle and Old Egyptian forms. The Hieroglyphica catalogue is reproduced online in [Encoding Egyptian Hieroglyphs in Plane 1 of the UCS](http://std.dkuug.dk/JTC1/SC2/WG2/docs/n1944.pdf) [http://std.dkuug.dk/JTC1/SC2/WG2/docs/n1944.pdf] (Michael Everson, 1999) a key reference relating to the proposed inclusion of Egyptian Hieroglyphs in the Unicode Standard (see www.unicode.org). A later revision of *Hieroglyphica* (2000) presents drawings of about 6,500 signs.

Meanwhile, the treatment of modern scripts on computers has become based on the Unicode standard. Unicode 1.0 was published in 1992 and is currently in version 4.0 (2003). The standard is subject to an ongoing process of refinement while maintaining compatibility with earlier versions. Unicode is the primary character coding standard for Windows, MacOS, Unix and Linux, although it should be noted that the implementations of these operating systems and the applications used on them tends to be some years behind the latest Unicode Standard. Recent standardisation activity in Unicode has included ancient scripts such as Linear B Syllabary and Ideograms, Ugaritic and Archaic Greek Musical Notation. Coptic has been disassociated from Greek to enable it to be treated as a script in its own right.

Beyond simple script representation as characters, XML (which uses Unicode as its character coding) has become strongly established and supported as the preferred notation for representing more complex structures and relationships.

OpenType (an evolved version of TrueType that also supports the Adobe tradition) is becoming established as the standard for fonts on modern computers (Windows 2000/XP, MacOS X, recent Linux and Unix systems). Notable is the extended support for joining cursive scripts such as modern Arabic. Egyptological software such as InScribe et al need Unicode standardisation and obsolescence of earlier computer systems before they can switch fully into OpenType but we hope this is near/medium future not a decade away. It is possible, even probable, that OpenType evolve to better support hieroglyphic text in this timeframe. We may even get to see colour text this decade! On a more practical note, we can reasonably expect technology more suitable for dealing with cursive hieroglyphs and hieratic. The [OpenType specification](http://www.microsoft.com/typography/otspec/default.htm) [http://www.microsoft.com/typography/otspec/default.htm] is highly technical!

It is likely that Unicode will include something Oxford/Gardiner sign list characters sometime in the next 2-3 years, a wider repertoire to follow if the additional signs can be put on a sound footing. The [Script Encoding Initiative](http://linguistics.berkeley.edu/sei/) [http://linguistics.berkeley.edu/sei/] is one attempt to encourage this process of enabling historical scripts, including ancient Egyptian, to be used with modern technology.

Meanwhile, a specification for coding hieroglyphs in the Unicode Private use zones is being developed - see [GameSet Resources: Unicode](#) [reproduced below].

Coding is no panacea for historical documents, especially for a language like ancient Egyptian whose history spans several millennia. Good quality facsimile and epigraphy play important roles and there are many elements of document analysis beyond phonetic and hieroglyphic transcription into canonical or idealised character representations. Nevertheless, much can be done through a straightforward XML/Unicode Egyptological standard to augment and modernise MdC and we hope to see this happen alongside any Unicode standardisation timeframe.

Overall a complex subject and barely touched in this note. However it is important to note that were technology not moving apace and we were in a static MdC universe, there is a huge amount of worthwhile exploration to be done in areas of database and publication of coded texts. Certainly new standards will open up new opportunities but there is no reason to defer many kinds of work. It can be safely assumed translation from current formats to whatever follows will be simple to automate on computer.

GameSet Resources: Unicode (February 2005, updated December 2005)

Introduction

The treatment of modern scripts on computers has become based firmly on the Unicode standard (see www.unicode.org). Unicode 1.0 was published in 1992 and is currently in version 4.0 (2003). The standard is subject to an ongoing process of refinement while maintaining compatibility with earlier versions. Unicode is the primary character coding standard for Windows, MacOS, Unix and Linux, although it should be noted that the implementations of these operating systems and the applications used on them tends to be some years behind the latest Unicode Standard.

Ancient scripts and Unicode

Ancient scripts are regarded as part of the eventual scope of Unicode although, understandably, the main focus thus far has been on developments relating to modern scripts. Recent standardisation activity in Unicode has included ancient scripts such as Linear B Syllabary and Ideograms, Ugaritic and Archaic Greek Musical Notation. Coptic has been disassociated from Greek to enable treatment

as a script in its own right. The [Script Encoding Initiative \(SEI\)](http://linguistics.berkeley.edu/sei/) [http://linguistics.berkeley.edu/sei/] is one attempt to encourage this process of enabling historical scripts to be used with modern technology, including complex scripts such as Ancient Egyptian.

Unicode is not a panacea for working with historical scripts and associated languages. Often, reproduction of sources by facsimile or representation through epigraphical considerations is more appropriate than transcription into a modern idealisation of the script. Notations can be used to express factors such as different usages of a script, its evolution, and linguistic considerations. Nevertheless, Unicode representations have clear applications and can yield a useful foundation for more elaborated treatments.

Egyptian Hieroglyphs and Unicode

Inclusion of Egyptian Hieroglyphs in Unicode has been a discussion point for around fifteen years. In 1998, a detailed proposal was written by Michael Everson - see [Encoding Egyptian Hieroglyphs in Plane 1 of the UCS](http://std.dkuug.dk/JTC1/SC2/WG2/docs/n1944.pdf) [http://std.dkuug.dk/JTC1/SC2/WG2/docs/n1944.pdf]. This proposal reproduced the Hieroglyphica (1993) catalogue of signs with some additional characters derived from the Manuel de Codage (MdC) notation. However the consensus seemed to be at that time a more suitable starting point would be a smaller set based on the Oxford/Gardiner hieroglyphic font. As further work was done on the script, additional characters could be added to the Unicode standard. A straightforward and practical proposal that is still pending.

The lack of a Unicode standard for Egyptian Hieroglyphs in recent years has been a problem for those involved in the subject, although much can be done in alternate or traditional coding schemes (such as MdC).

Egyptian Hieroglyphs in the Unicode Private Zone

Unicode Private Use Zones can be used for scripts not yet part of the standard. Some practical problems associated with using Egyptian Hieroglyphs can be addressed by working in these zones. Font organisation is the best known issue although this is just one application.

A proposed specification for such an Egyptian Private Zone (EGPZ) is being written by Bob Richmond (Saqqara Technology). Two zones are allocated:

- **BMP Zone** (Private Use Area 0xE000 to 0xF8FF). Commonly used signs, including the Oxford/Gardiner Set and extensions found in modern fonts (excluding the full CCER Extended Library). Accessible to older applications and operating systems that do not support surrogate pairs or full UTF-8. Transitional - over time a formal Unicode standard alongside declining significance of older applications will probably render this zone superfluous.
- **Plane 15 Zone** (Extended Private Zone A 0xF0000 to 0xFFFFFD). A catalogue of all reported signs given in fonts and sign lists, including composite forms and variations that would not normally count as Unicode characters. Specialist in nature but useful. Wider in scope than an anticipated formal Unicode standard. Version 1.0 essentially Gardiner/Oxford plus CCER Extended Library plus additional requests.

A detailed draft of this specification is being prepared for publication here on GameSet. There is an opportunity for comment and feedback before a 1.0 release is published (probably in Spring 2006 - Bob is also involved with work on a formal Unicode specification and would like the larger EGPZ scope to fit with this.). Although EGPZ has no 'official' status, it is hoped this EGPZ will provide a useful stepping stone to true Unicode applications when the formal Unicode specification is agreed. See [GameSet Resources: Unicode: Egyptian Hieroglyphs in the Private Zone](#) [reproduced below]. or contact saqqara@csi.com with comments.

GameSet Resources: Unicode: Egyptian Hieroglyphs in the Private Zones (February 2005, updated December 2005)

Introduction

The Egyptian Hieroglyphic script has been discussed as an addition to the Unicode standard for character encoding since the inception of Unicode. See www.unicode.org). There are indications that progress may be made in the next year or two but it is likely to be some years before such a standard is approved and useable.

The Unicode standard has private zones which can be used for (as yet) unstandardized scripts. As far as the author is aware, no other proposal for implementation has been published to date. This document therefore outlines a proposal for Egyptian Hieroglyphs in the Private Zone (EGPZ). This informal standard has some practical benefits in the construction of software and data. It is intended to help extend the scope of applications working with the ancient egyptian script.

There are definite limits as what can be accomplished in the Private Zone. Simple actions such as display of hieroglyphs in a word processor document can be accomplished by choosing an EGPZ-compatible font but formatting and other actions that depend on knowing what a character represents remain limited to specialist hieroglyphic software designed to understand EGPZ. For that reason, this specification is aimed largely at software writers working on programs that work with hieroglyphs to deliver the benefits to end users that derive from a consistent scheme.

Nevertheless, if you are a user of hieroglyphic software, you may be interested in the repertoire of signs under consideration and the potential of the system.

The EGPZ proposal is still being developed and is subject to change and clarification.

Design considerations

- **Scope of a Unicode approach to hieroglyphs.** EGPZ does not make any statement about when it is, or is not, appropriate to use Unicode to represent texts written in the hieroglyphic script. It simply allows coding of current practice, drawn from the tradition of hieroglyphic fonts and conventions used in contemporary computer software.
- **Target platforms.** EGPZ should be viable on Linux, Unix, MacOS and Windows machines with applications and development tools that support Unicode. Where this implies some restrictions, popularity should be taken into account. Examples. The old Windows 98 platform is far more widely used at present than Linux. Many applications and software products still in use do not support Unicode outside the BMP through surrogate pairs or full UTF-8.
- **Sign list inclusivity.** Unicode is about characters, not glyphs. This is a complex issue with the hieroglyphic script. Factors include: the number of graphically distinct sign forms, the long historical context and the fact that applications are often descriptive (of ancient texts) rather than productive (origination of new texts). EGPZ sidesteps this issue by allowing all modern representations to be in scope, including rare variants and combinations of signs. A formal Unicode standard would likely take a different approach.
- **Compatibility with other coding schemes.** All signs will be documented with a Gardiner-style code where possible and this will normally correspond to the code published in

Hieroglyphica (2000). Exceptions follow the convention used by InScribe 2004 which assigns codes to Gardiner/Oxford signs not given in *Hieroglyphica* .

- **Arrangements of hieroglyphs.** At version 1.0, there are no control codes analogous to the MdC stacking (':') and juxtaposition ('*') codes or similar - such is regarded as the province of 'generic' Unicode features or higher-level markup or protocols. This does not preclude the possibility of such elements in a future revision (or indeed in the Unicode Standard itself) if experimental evidence for a requirement can be presented.
- **Implementations.** EGPZ provides a consistent approach to some elements of working with the hieroglyphic script. Implementations are expected to make different decisions about how to use the system. Examples. 1. It is likely that early implementations operate in the BMP and not the Extended (Plane 15) zone. 2. A software approach may take the point of view that treating sign combinations as distinct codes is unhelpful and this ought to be done at a higher level of description than Unicode. This is fine so long as the basic defined set of hieroglyphs is supported.
- **Audience.** Specifications should be written for use by software developers but the inventory of signs and codes should be accessible to Egyptologists.

Egyptian Hieroglyphs in the BMP

Space for 6400 codes is available in the Private Use Area (PUA) of the Unicode Basic Multilingual Plane (BMP) in the region 0x0E000 to 0x0F8FF. EGPZ partitions this as follows:

- 0x0E000 to 0x0E7FF. Space for up to 2048 signs. The complete Oxford/Gardiner set (around 1000 signs) plus all signs found in other hieroglyphic fonts currently used on Personal Computers, except for the Extended Library from CCER. Useful characters not already in Unicode such as cartouche, serekh, Hwt start and end along with quadrant shading characters (as used in MdC) will also be defined in this block *Any information on relevant fonts or desirable signs would be appreciated, the block can be extended if necessary.*
- 0x0E800 to 0x0F3FF. Space for up to 3072 signs. Reserved for future revisions of the specification.
- 0x0F400 to 0x0F8FF. Space for 1280 signs. This is available for implementation-dependent features.

Formatting codes such as the ':', '*' and '&' operators found in MdC are **not** provided in version 1.0 since it is not clear how best to work with sign arrangements in a way that might obtain a consensus approach.

Programs and fonts (or graphic lists) that implement EGPZ/BMP are only required to implement a basic subset from the sign list given in *Egyptian Grammar* to be EGPZ-compliant. The other signs are optional.

The formal Unicode standard for Hieroglyphs, when available, is in Unicode Plane 1, not the BMP. Therefore, this part of EGPZ will still be required for compatibility with software that remains limited to the BMP. A transition period of several years can be anticipated and software should take a transition period into account.

A draft of the EGPZ/BMP specification is available in [EGPZ-BMP-0.1.pdf](#) [See www.egpz.com]. *This is fairly rough and published early for comment. The current BMP list numbers about 1400 signs with a further 300 (mainly Middle Egyptian) possible candidates for version 1.0. Additional suggestions are welcome.*

Egyptian Hieroglyphs in the Unicode Extended Private Zone A

Space for 65534 codes is available in the Unicode Extended Private Zone A (0xF0000 to 0xFFFFD) (also known as Plane 15 or EGPZ/EA).

EGPZ partitions this space as follows:

- 0xF0000 to 0xF1FFF. Space for up to 8K signs. At version 1.0, this is the Oxford/Gardiner Set merged with Hieroglyphica 2000 and any additional signs arising in the BMP implementation (including cartouch, shading etc.) or specifically requested. All these signs are documented with Gardiner-style codes following *Hieroglyphica* where appropriate. This section is essentially a catalogue of all signs used by current software.
- 0xF2000 to 0xFCFFF. Space for 40K signs. Reserved for future EGPZ expansion.
- 0xFC000 to 0xFDFFF. Space for up to 8K signs. Provided for private use. For example, if a program allows a user to define their own 'new' sign or sign combination, an (arbitrary) code value can be assigned from this range. Normally, this feature will be used together with some higher level protocol that allows a description of what the code means.
- 0xFE000 to 0xFFFFD. Space for up to 8K-2 signs. Provided for software implementation-specific private use. For example, a given program may want to use a sign variant or combination of signs not defined by EGPZ as part of its implementation. However, it should not use codes here for data exchange with other programs - the previous 8K range should be used in this case.

The codes defined here are simply meant to function as a complete catalogue of signs. The unused space of 40K is far larger than one would anticipate being needed but there appears no purpose in not reserving it.

Programs and fonts (or graphic lists) that implement EGPZ/EA are only required to implement the same basic subset from the sign list given in *Egyptian Grammar* as used in EGPZ/BMP. The 7000+ codes defined at version 1.0 are there to be used if needed but implementations are encouraged to implement only those codes that are needed for a given purpose. For instance Hieroglyphica defines many signs and variants from the Greco-Roman periods that are not useful for Old, Middle or Late Egyptian work.

An early draft of the EGPZ/EA specification is available in [EGPZ-EA-0.1.pdf](#) [See www.egpz.com]. *Comments and other feedback appreciated.*

Feedback

A detailed draft of this specification is being prepared for publication here on GameSet. The early specifications given here in PDF are being adjusted to fit better with possible directions to a formal Unicode standard for Egyptian. There is opportunity for comment and feedback before a 1.0 release is published (probably in Spring 2006). Although EGPZ has no 'official' status, implementations are expected this year and it is hoped this EGPZ will provide a useful stepping stone to true Unicode applications when the formal Unicode specification is agreed. Contact saqqara@csi.com with comments or for further details.

Bob Richmond
Saqqara Technology Ltd.