Hieroglyphs in Unicode

New Developments

Debbie Anderson, Script Encoding Initiative, Dept. of Linguistics, UC Berkeley; http://linguistics.Berkeley.edu/sei; dwanders@berkeley.edu and Barbara Richter, Dept. of Near Eastern Studies, UC Berkeley

ARCE 67th Annual Meeting, April 16 2016

1. What is Unicode?

- Unicode is the standard that makes it possible to type, send, display, and store text electronically on modern computers and mobile devices.

- Unicode also serves as the standard followed by search engines, fonts and keyboards.

- If your computer or phone is missing a font, or if the characters are not in Unicode, you may see:

□□□□□□□□□□□□   ������   Åë∎é‖÷ëù¬Å

2. Why is Unicode important to Egyptologists?

- Will enable Egyptologists to type, send, and search texts on any platform (Mac, PC, Linux) with most off-the-shelf software

- Can help save time, otherwise spent on creating text, and converting to various formats

- As an international standard, won't become obsolete

3. Background

- First preliminary proposal in **1997** by M Everson, with several revisions

- Unicode published the Gardiner set (with supplements) in **2009** (Unicode 5.2)

- No other proposals until drafts for Extended Egyptian appeared in **2015** and **2016** by M Suignard

1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020 2022

First prop.

Rev. prop.

Rev. prop.

Final Prop.

Published in Unicode

New proposal

4. Current Situation

- Text-processing editors
  - JSesh and VectorOffice
  - WinGlyph /MacScribe no longer supported
  - create images, which can't be searched
- Unicode characters
  - plain text – not images
  - displayable with free fonts (Noto Sans Egyptian Hieroglyphs, Aegyptus, NewGardiner)
  - limited to 1071 characters (Gardiner set plus supplements), but they are searchable
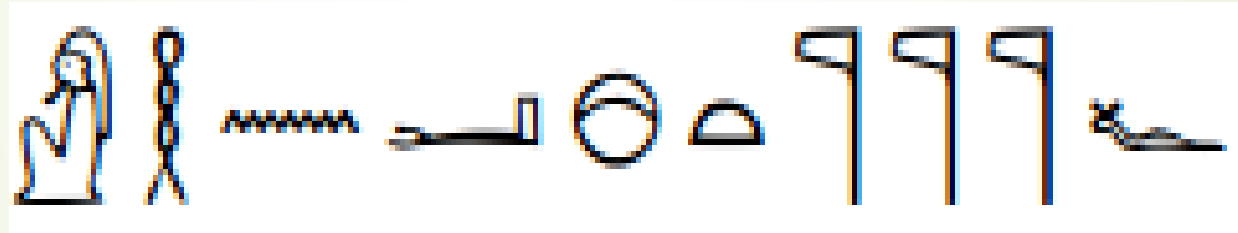
4. Current Situation

- Can search characters in Unicode on Chrome, Firefox and MS Word (below is a search for 🦌)
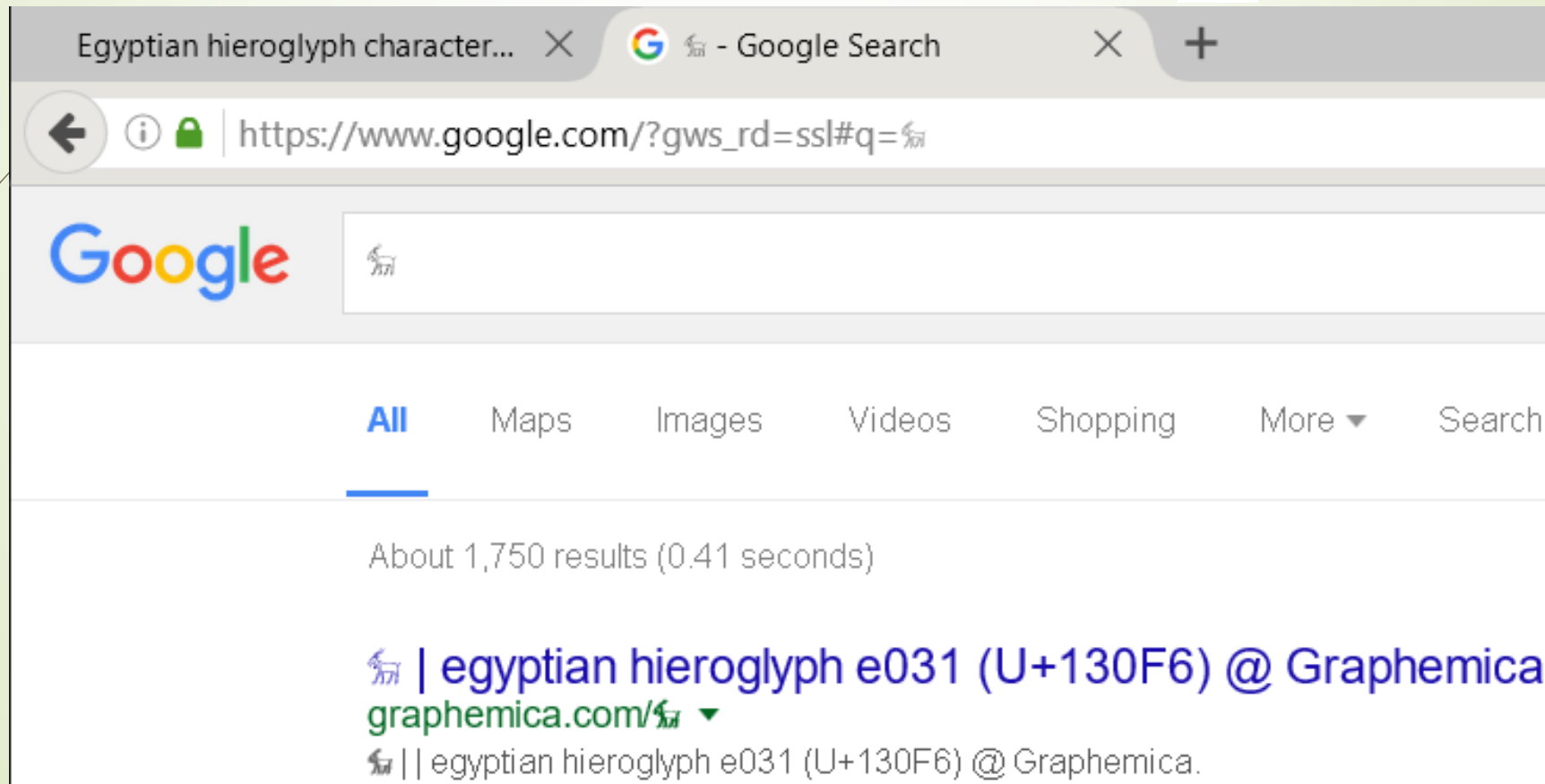
4. Current Situation

- Problem: Unicode characters can only be laid out linearly, so you can't get desired layout:

- Users were left to rely on specialized software that created images with clusters and additional signs

5. Developments:
a) Three new control characters

- Will allow clustering with Unicode characters in text:

- Won't require special software; uses * : and + (="&" in MdC)

- On track for future Unicode version (2017?)

- Planned support in Windows 10 (MS Office) and other platforms

- Software being developed that will allow use of these characters, and will convert JSesh > Unicode; see see: http://hieroglyphseverywhere.blogspot.co.uk/

5. Developments:
b) Unicode Character Picker

- Various capabilities; one way to make Unicode characters more accessible

- Freely available at http://r12a.github.io/pickers/egyptian/

- Created by Richard Ishida; to suggest additional features or send comments, write to: ishida@w3.org

text area

Show codepoints Convert to escapes MdC to Hieroglyphs Hieroglyphs to MdC MdC to WikiHiero Change Transliteration Phonetic look up Make example Character markup

A. Man and his occupations (man seated)

Autofocus: On Off

D L C L E

SPACE   NBSP   RLO   PDF   ⊂ ⊐ < > * _ !

Search for: keywords   X

Font list: Noto Sans Egyptian Hieroglyphs WF ▼   Custom font: Noto Sans Egyptian Hierogly   Set dimensions: 45 3 1.5   Add codepoint: Search for: Catalog numbers   X   Help ?
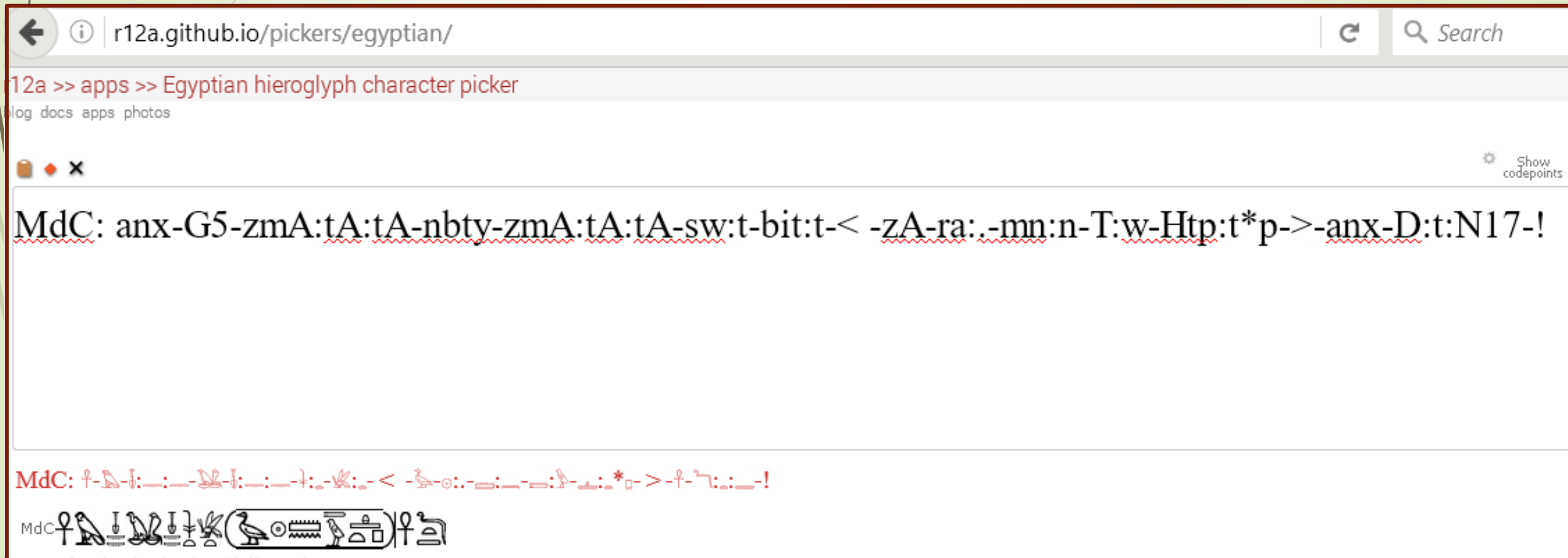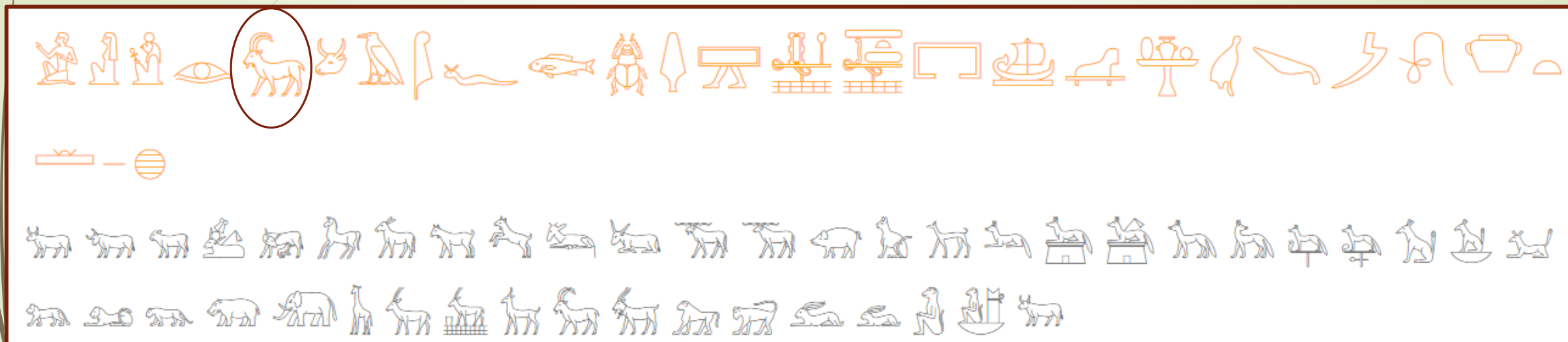
5. Unicode Character Picker

Features: Can change MdC to hieroglyphs
(Unicode characters appear in-line with "MdC to
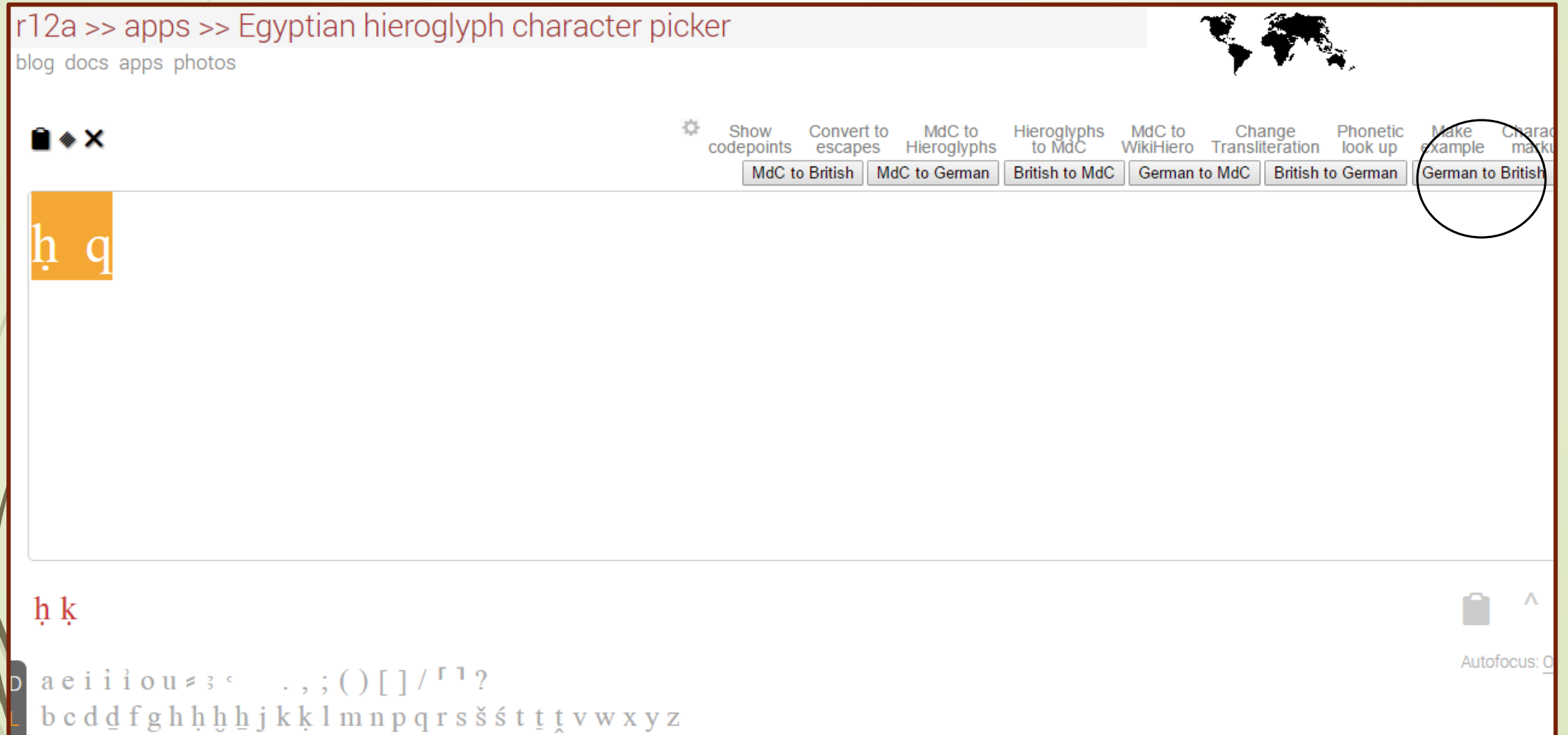**Hieroglyphs"option or as images if you select "MdC to WikiHiero")**

5. Unicode Character Picker

Features: Can show all "mammals" currently in Unicode Egyptian hieroglyphs

5. Unicode Character Picker

Features: Can change transliteration: From "German" to "British", etc.

5. Unicode Character Picker

Features: Can change hieroglyphs to MdC

5. Developments:
c) Extended Egyptian in Unicode?

2016 Proposal by Michel Suignard to extend Egyptian

- Total 6269 characters; need careful review
- Based on Hieroglyphica and H Extended
  - Glyph-based list with many variants
  - Widely used (JSesh, VectorOffice, Aegyptus, Winglyph, TLA)
  - Lacks source documentation
- Includes additions from Aegyptus font
- Stepping stone towards future proposal

6. Next Steps: Review of proposal

1. Variants

Problem

Unicode character: N08

Hieroglyphica distinguishes between N08 and

N8A and N8D

Question
Are N8A and N8B separate characters or glyph variants?

Need to decide what to do

6. Next Steps : Review of proposal

Unicode N08   N8A   N8D

Character or not?

- If there is a **systematic** distinction between N8A, N8D and N08, it would be good to retain semantic distinction (and encode them separately)

- If there is **no systematic distinction semantically** but Hieroglyphica N8A and N8D only reflecteda form from one source, then it may be a **stylistic** distinction

- Or, if one or both are glyph variants for which the distinction **needs to be kept for palaeographic purposes**, they could be eligible to be separate characters

6. Next Steps: Review of proposal

Example

- First four ( ) are typographical styles of small letter s (italic s, script s, and bold s)
  - reflect a stylistic distinction, which can be kept in fonts
- Historic form of s (long s,)
  - a distinction no longer kept today (except in Gaelic and Fraktur), but maintained in historical materials, so it has been encoded as a separate character

6. Next Steps: Review of proposal

If the distinction is only stylistic

- If it is *minor*, such as a font design, do nothing, but can put glyph into font (no separate character)

- If the distinction is *systematic* across the entire corpus, but varies across certain materials, for ex., a "Variation Selector" might be appropriate

- If the distinction is a *significant paleographical distinction* (such as long s), it may need to be separately encoded

6. Next Steps: Review of proposal

Need to decide:

- Which characters should be separate characters
- Which signs in the proposal are just stylistic variants (so could be included in a font, but not needed as separate characters)
- Which signs reflect important distinctions for palaeography that need to be kept in plain text
- Which distinctions are found systematically across the corpus, and need to keep distinction in plain text (and might be eligible for use with Variation Selector)

6. Next Steps: Review of proposal

2. Glyph changes (page 5 of proposal)

3. Identify sources for Hieroglyphica characters

- Make information openly accessible

- Develop database/lookup tool

Sample prototype

Unicode Egyptian Database Lookup

Lookup | 13052

About the Unicode Egyptian Database Lookup Tool

As a handy reference, the Unicode Consortium here provides a search interface to the Unicode Egyptian Database.

The Unicode Egyptian Database organizes information relating to the properties of Egyptian hieroglyphs. Documentation is available in TBD.

For production reasons, the version of the Unicode Egyptian Database available via this lookup tool may not yet be in sync with the latest version of the Unicode Standard. For access to the most recent version of the raw data files, see http://www.unicode.org/Public/UCD/latest/.

The lookup interface on this page provides access to individual hieroglyphs through the "Lookup" button and text field above. Enter the four- or five-digit hexadecimal identifier for the character (if you know it), or copy and paste a character (if you have one), and then click the "Lookup" button.

The resulting data set will contain various types of information available in the Unicode Egyptian Database, for example, mappings to legacy catalog identifiers, references to major dictionaries, and meaning and pronunciation information according to various authorities.

TBD: Augment this according to whatever additional lookup features are added.

Disclaimers

The Unicode Egyptian Database is provided as a public service by Unicode, Inc. These data are provided as-is by Unicode, Inc. (The Unicode Consortium). No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided.

The data in the Unicode Egyptian Database derives from various sources, as documented in TBD.

Sample prototype

Unicode Egyptian Data for U+13052

| Lookup | | | |

Glyphs

The Unicode Standard	Your Browser

Encoding Forms

Decimal	UTF-8	UTF-16	UTF-32
77906	F0 93 81 92	D80C DC52	00013052

Catalog Information

Catalog	Value
Primary catalog number	HB003
IFAO	54,5
Cauville	B,28,4
Gardiner	B3
Kurth	1,137,51
Val Phon	81,27

Sample prototype

Glyphs

The Unicode Standard	Your Browser

Hieroglyphica	B3
Hieroglyphica extended	
Unikemet	B003
Douros	

Other Data

Data type	Value
Description	woman giving birth
Components	seated woman, baby
Determinative	give birth
Logograph	
Phoneme	
Variants	

Sample prototype: New characters

Unicode Egyptian Data for U+13B62

Lookup

Glyphs

The Unicode Standard	Your Browser

Encoding Forms

Decimal	UTF-8	UTF-16	UTF-32
80738	F0 93 AD A2	D80C DF62	00013B62

Catalog Information

Catalog	Value
Primary catalog number	HD197
IFAO	107,5
Cauville	D,65,4
Gardiner	
Kurth	2,172,52
Val Phon	168,439
Hieroglyphica	D197

6. Next Steps: Review of proposal

4. . Come up with practical guidelines:

- What if no attestations for the sign can be found?
- What if no meaning can be identified?

Note: The goal of Unicode is not to capture fine palaeographic detail, but arrive at a standardized way to represent Egyptian hierogl. text in databases, email, the web, documents, etc.
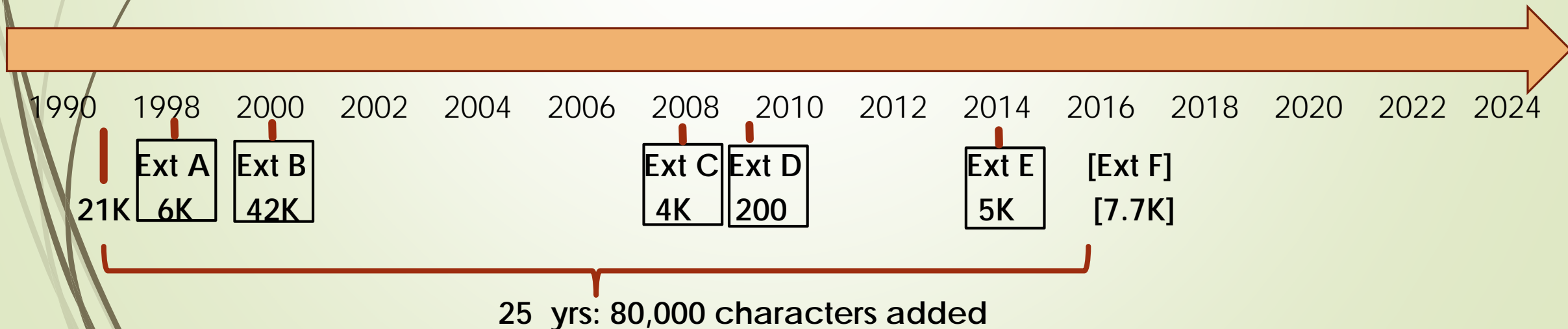
6. Next Steps

Timeframe

- Ongoing work
  - in Liège/Berlin est. to take 4+ years
  - Independent effort by D. Meeks
- In the meantime, scholars will be limited in ability to search, send text, etc. for many years
- Once a proposal has general user consensus, it takes about 2 years to go through the standards process

Note: Incremental additions to Unicode can be made

6. Next Steps
Compare CJK (Chinese Japanese Korean)

- Took existing standards and developed working implementations of them

- Experts worked on next chunk

- Extensions include modern additions, simplifications, vast number of historical materials

1990   1998   2000   2002   2004   2006   2008   2010   2012   2014   2016   2018   2020   2022   2024

| Ext A | Ext B | | Ext C | Ext D | | Ext E | [Ext F] |

21K 6K   42K   4K   200   5K   [7.7K]

25 yrs: 80,000 characters added

6. Next Steps
Comments

- Getting characters into Unicode need not be an all-or-nothing proposition

- Unicode should not be viewed as a definitive dictionary of Egyptian hieroglyphs

- Gardiner set was included in Unicode as it was an agreed-upon set

- Can add characters on which there is consensus

- Can remove any problematical cases for future study

6. Next Steps

- Meeting in Cambridge in July 2016 (organized by N Strudwick)

- Arrange other meetings to review proposal and decide on handling variants?

7. The future

- Development better input mechanisms
- Creation of Winglyph to Unicode converter?
- Follow up with implementers and font providers to support mirroring and vertical text
- Decide on agreed-upon character to transliterate yod

7. The future

- New developments

- With Unicode, Egyptologists are poised to be able to speed up publication, increase their potential research capabilities (with search), and have better support from available software , but help is needed from you

Questions? Comments?

Special thanks to:

Barbara Richter

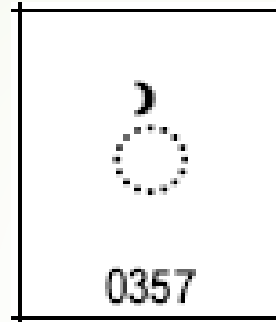Ken Whistler

Michel Suignard

Richard Ishida

Bob Richmond

Addendum: Representing yod

Three options available: use lowercase letter i and uppercase I with one of the following:
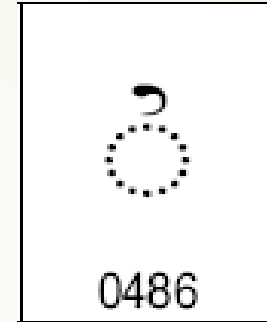
0313 is generic comma above, intended for Greek

0357 is half-ring diacritic for Uralic Phon Alph, used Semitic transliteration

0486 is historic Cyrillic variant of Greek psili pneumata