A. Administrative

1. Title: Proposal to add Greek epigraphical characters to the UCS

2. Requester's name: Nick Nicholas

3. Requester type: Expert contribution

4. Submission date: 2005–05–22

5. Requester's reference: —

6a. Completion: This is a complete proposal

6b. More information to be provided? No.

B. Technical—General

1a. New Script? Name? No.

1b. Addition of character(s) to existing block? Name? Yes. Greek or Greek Extended.

2. Number of characters in proposal: Six

3. Proposed category: B.1. Specialized (small collections of characters)

4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): Level 1 noncombining character

Is a rationale provided for the choice? No

5. Is a repertoire including character names provided? Yes

a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? Yes

b. Are the character shapes attached in a legible form suitable for review? Yes

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? David Perry

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: —

7. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? Yes

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? Yes

8. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? Yes

9. Additional Information:

C. Technical—Justification

1. Has this proposal for addition of character(s) been submitted before? Yes, as *L2/05-002 Proposal to add Greek Letter Lowercase Heta and Greek Letter Capital Heta* (supercedes L2/04-388), and *L2/05-003 Proposal to add Greek epigraphical letters* (see also L2/04-389).

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? Yes

If YES, with whom? Subscribers of Unicode Greek and Epigraphical mailing lists.

If YES, available relevant documents: Feedback obtained from specialists to Nick Nicholas and Deborah Anderson available on request.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Scholarly community and individuals interested in Greek linguistics and epigraphy

Reference: —

4. The context of use for the proposed characters (type of use; common or rare): Common in Greek epigraphy, occasional in Ancient Greek linguistics

Reference: —

5. Are the proposed characters in current use by the user community? Yes

If YES, where? Reference: Characters (or corresponding codepoints in transliteration schemes) are present in various publications on Ancient Greek linguistics, and in publications and digitisations of epigraphic corpora

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? Yes

If YES, is a rationale provided? Contemporary use, keeping character together with other Greek characters

If YES, reference: —

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? No.

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? No (but see arguments in previous submissions)

If YES, is a rationale for its inclusion provided? —

If YES, reference: —

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No

If YES, is a rationale for its inclusion provided? —

If YES, reference: —

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? Yes

If YES, is a rationale for its inclusion provided? See previous submissions.

If YES, reference: —

11. Does the proposal include use of combining characters and/or use of composite sequences? No

If YES, is a rationale for such use provided? —

If YES, reference: —

Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? —

If YES, reference: —

12. Does the proposal contain characters with any special properties such as control function or similar semantics? No

If YES, describe in detail (include attachment if necessary) —

13. Does the proposal contain any Ideographic compatibility character(s)? No

If YES, is the equivalent corresponding unified ideographic character(s) identified? —

If YES, reference: —

Proposal

I propose the following six characters for inclusion in the Universal Character Set:

U+0370	GREEK CAPITAL LETTER HETA	⊢
U+0371	GREEK SMALL LETTER HETA	⊢
U+0372	GREEK CAPITAL LETTER ARCHAIC SAMPI	⊓
U+0373	GREEK SMALL LETTER ARCHAIC SAMPI	⊓
U+0376	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA[1]	И
U+0377	GREEK SMALL LETTER PAMPHYLIAN DIGAMMA	и

Issues involving these characters have been discussed extensively in previous submissions (L2/05-002 for Heta, L2/05-003 for the other epigraphical characters), and are not reiterated here. Likewise examples of the use of characters in print are not repeated. In the

---

[1]Renamed from *Greek Letter Tsan Or Pamphylian Digamma* in previous proposals. The Pamphylian use of the glyph is clearly more productive than the Arcadian, which is limited to one document—however important that usage may be for Greek historical phonology. Moreover, as pointed out to me by Peter Haarer, the name I have devised, *Tsan*, presupposes an uncertain connection of the Arcadian letter to San.

light of feedback received since, however, I note the following:

- There is not a strong tradition of casing any of these characters other than heta; case has rather been introduced to anticipate the possibility that it will be required, as has already taken place for Greek numerals.

- The proposed codepoints are intended to err on the side of unification rather than disunification. They thus are to encompass either disparate glyphs with the same or similar underlying phonetic value, or identical glyphs representing distinct phonetic values in quite separate dialects. While there is a possibility that disunification will be called for in the future (and some respondents have said as much), the conflations proposed are expedient. In particular:

  - *Heta* is to encompass the various epichoric letters for /h/ when encoded as letters rather than the diacritic *Dasia* (U+0314 Combining Reversed Comma Above). The reference glyph is the 'tack', for the sake of distinctiveness (although it is identical to the non-alphabetic reference glyph for U+10142 Attic Acrophonic Symbol One Drachma), However the dominant means of encoding this letter has long been the Latin glyph <h>.

    Several respondents (Elaine Matthews, R.J.E. Thompson, John Mansfield, Peter Haarer) have indicated that they would desire a distinct Greek codepoint even if they continued to use the Latin glyph, and that their current encodings already make such a distinction. A heta codepoint allows such a distinction to be made, and for the tack and <h> to be treated as glyph variants. The distinct codepoint also allows Unicode to be agnostic in unresolved debates on the proper encoding of heta (whether as <h> or tack, whether a Greek <h> should be differentiated from the Latin <h>, whether <H> should be allowed for capital heta, etc.)

  - The *archaic sampi* unifies all Ionian glyphs used with a phonetic value of /ss/ or similar, recognising that the identification of these glyphs with the numeral sampi is tentative. (The older name *disigma* could be used to disambiguate the numeric from the alphabetic use, but this seems unnecessary.) The Pamphylian psi-like letter used for /s/, /ss/, /ps/ is also intended to be represented by this codepoint as a matter of expediency, although it is not certain whether it shares a common pedigree with the Ionian letter.

  - Pamphylian digamma is intended for the digamma variant in the dialect transliterated as <v> and believed to have had the value [w]. As a matter of expediency, I suggest it also be used for the linguistically important but unrelated one-off use in Arcadia of the same glyph to represent /ts/ (which I have christened 'Tsan'). John Mansfield found the conflation surprising, and the characters do appear unrelated; but two distinct codepoints for an identical glyph, one of them appearing in only one document, seems to me untenable.

Glyphs

As noted, heta encompasses the glyphs tack, Latin <h>, and boxed heta; the tack is chosen as a reference glyph for its distinctiveness.

The reference glyph for archaic sampi should accentuate the letter's 'serifs', to avoid confusion with capital tau (despite their frequent conflation in heritage data).

Sorting

Heta should be encoded as a letter, and it should sort either immediately after or immediately before eta; since the characters were almost always mutually exclusive (with the exceptions of Delphi, Heraclea/Tarentum, and Cnidus), there is no established ordering between them. (Peter Haarer has suggested heta come first, since eta is derived from heta, and so "could perhaps be described as a derivative".) Most indexes containing heta (e.g. Buck 1955) give it secondary weighting, as if the heta was a rough breathing in conventional orthography; such indexes also ignore digamma, so that the words are sorted as if they are in standard Greek orthography (digamma was dropped in Attic, the classical standard). Such weighting should not be enforced in the Default Collation, however, and the sorting of heta as a distinct letter is found in histories of the Greek script like Jeffery's.

The archaic sampi should sort with the numerical sampi, just as the archaic koppa sorts with the numerical koppa.

The default sorting location of Pamphylian digamma should be with normal digamma. If the codepoint is used to represent tsan, and needs to be sorted with, say, san, this would be a matter of introducing an Arcadian-specific sorting order, and possibly having words sort in two different locations depending on language markup. (The situation would be akin to a single index containing Swedish and German words with <ö>, the German instances sorting after <o>, and the Swedish after <z>.)

Linebreaking and combinatorics

All characters in the proposal are used like any other Greek alphabetic character. All characters can appear at the beginning or a middle of a Greek word. (In the case of heta, this applies to dialectal/early Greek, although the equivalent rough breathing diacritic does not appear word-medially in canonical orthography.) In grammatical Greek text, the characters would not appear word-finally, but the incomplete words characteristic of epigraphy make this possible (e.g. Τεⱶ[ ). The characters combine with the diacritics endemic to epigraphy—notably dot below.

Properties

UCD Entry

```
aaaa;GREEK SMALL LETTER ARCHAIC SAMPI;Ll;0;L;;;;;N;;;bbbb;;
bbbb;GREEK CAPITAL LETTER ARCHAIC SAMPI;Lu;0;L;;;;;N;;;;aaaa;
```

cccc;GREEK SMALL LETTER PAMPHYLIAN DIGAMMA;Ll;0;L;;;;;N;;;ffff;;

dddd;GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA;Lu;0;L;;;;;N;;;;eeee;

eeee;GREEK SMALL LETTER HETA;Ll;0;L;;;;;N;;;bbbb;;

ffff;GREEK CAPITAL LETTER HETA;Lu;0;L;;;;;N;;;;aaaa;

DUCET Entry

03E1 ; [.110A.0020.0002.03E1] # GREEK SMALL LETTER SAMPI

03E0   ; [.110A.0020.0008.03E0] # GREEK LETTER SAMPI # same distinction made between Koppa and Archaic Koppa

aaaa ; [.110B.0020.0002.03E1] # GREEK SMALL LETTER ARCHAIC SAMPI

bbbb ; [.110B.0020.0008.03E0] # GREEK CAPITAL LETTER ARCHAIC SAMPI # same distinction made between Koppa and Archaic Koppa

03F8 ; [.110C.0020.0002.03F8] # GREEK SMALL LETTER SHO

03F7 ; [.110C.0020.0008.03F7] # GREEK CAPITAL LETTER SHO

03DD ; [.10EE.0020.0002.03DD] # GREEK SMALL LETTER DIGAMMA

cccc   ; [.10EE.0020.0002.eeee] # GREEK SMALL LETTER PAMPHYLIAN DIGAMMA

03DC ; [.10EE.0020.0008.03DC] # GREEK LETTER DIGAMMA

dddd   ; [.10EE.0020.0008.ffff] # GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA

03DB ; [.10EF.0020.0002.03DB] # GREEK SMALL LETTER STIGMA

03DA ; [.10EF.0020.0008.03DA] # GREEK LETTER STIGMA

eeee ; [.10F1.0020.0002.gggg] # GREEK SMALL LETTER HETA

03B7 ; [.10F1.0020.0002.03B7] # GREEK SMALL LETTER ETA

1D6C8 ; [.10F1.0020.0005.1D6C8] # MATHEMATICAL BOLD SMALL ETA; QQK

1D702 ; [.10F1.0020.0005.1D702] # MATHEMATICAL ITALIC SMALL ETA; QQK

1D73C ; [.10F1.0020.0005.1D73C] # MATHEMATICAL BOLD ITALIC SMALL ETA; QQK

1D776 ; [.10F1.0020.0005.1D776] # MATHEMATICAL SANS-SERIF BOLD SMALL ETA; QQK

1D7B0 ; [.10F1.0020.0005.1D7B0] # MATHEMATICAL SANS-SERIF BOLD ITALIC SMALL ETA; QQK

ffff ; [.10F1.0020.0008.hhhh] # GREEK CAPITAL LETTER HETA

0397 ; [.10F1.0020.0008.0397] # GREEK CAPITAL LETTER ETA

1D6AE ; [.10F1.0020.000B.1D6AE] # MATHEMATICAL BOLD CAPITAL ETA; QQK

1D6E8 ; [.10F1.0020.000B.1D6E8] # MATHEMATICAL ITALIC CAPITAL ETA;

QQK

1D722 ; [.10F1.0020.000B.1D722] # MATHEMATICAL BOLD ITALIC CAPITAL ETA; QQK

1D75C ; [.10F1.0020.000B.1D75C] # MATHEMATICAL SANS-SERIF BOLD CAPITAL ETA; QQK

1D796 ; [.10F1.0020.000B.1D796] # MATHEMATICAL SANS-SERIF BOLD ITALIC CAPITAL ETA; QQK