

# ISO/IEC JTC1/SC2/WG2 N 2168

Author: Kent Karlsson  
Status: Expert Contribution  
Date: 2000-03-02

## Comments on 'Math Alphanumeric' characters for 10646-2.

These are some a bit more detailed comments supplementing the Swedish NB NO vote to the CD1 ballot on 10646-2. (See also N2169 on 'Language Tag' characters.)

The “math alphanumeric” characters (preformatted letters and digits [bold, fraktur, open-face, ...]) are suggested to be allocated in the plane 1.

The Swedish comment is to remove all text and tables referring ‘math alphanumeric’ characters (that are not already, by mistake, allocated in the BMP). The presence of the ‘math alphanumeric symbols’ characters (in plane 1) is one of the reasons for the Swedish NO vote on the 10646-2 CD.

This document gives motivations for the NO vote, motivations that for brevity are not included in the vote itself.

### a. History

There are some characters already in the UCS that are preformatted (except for size): “BLACK-LETTER...” (for fraktur), “DOUBLE-STRUCK...” (for open-face), and “SCRIPT...” letters. They cannot be removed now, unfortunately. But their presence do not constitute a precedent for including even more such characters. Their inclusion in the UCS should be regarded as a historic mistake, not to be repeated. The suggestion to include “math alphanumeric symbols” appears to stem from the MathML effort. However, the already existing preformatted characters should not be used, at all, in any kind of mark-up for math, as explained below (see points d-h).

### b. Acknowledgement of semantic distinction of certain font selections in math

It is true that mathematical expressions often use *certain* font distinctions to convey semantic distinctions. That, however, does NOT imply that it is appropriate to encode preformatted alphanumeric characters for this, as explained below. Nor does it mean that just any font change conveys such distinctions. Out of tens of thousands of fonts, only a handful of distinctions are recognised as meaningful distinctions for math expressions.

### c. Letter restrictions, multi-letter identifiers, and internationalisation

Some mathematical expressions use identifiers similar to those in programming languages. This is especially popular in computing science. The current proposal for math “alphanumeric symbols” covers essentially only A-Z (and basic Greek), while programming language identifiers are being generalised to allow for any written “word” in a natural

language. Why accept the “math alphanumeric symbols” with their limitations, when the rest of the computing world is being internationalised? See below for concrete proposals on what should be done for math instead (which does not involve allocating any new characters at all).

d. TeX/LaTeX, Omega/Lambda

TeX, or rather LaTeX, which is a macro package on top of TeX, is the today most widely used typesetting system for math used by mathematicians world-wide, when they typeset their own papers. (TeX/LaTeX are being generalised to Unicode/10646 in the Omega and Lambda efforts.) TeX uses commands (compare mark-up) to format the text, including math expressions. LaTeX has several commands like `\mathcal` that takes an argument with ordinary letters, and displays/prints these letters in a (pre-selected) script font. E.g. `\mathcal{ABC}` displays/prints ABC in the (pre-selected, math adapted) script font. Likewise `\frac{ABC}` displays/prints ABC in the (pre-selected) fraktur font. **These systems have no need whatsoever of any preformatted “math alphanumeric symbols” at separate code points from ordinary alphanumeric characters.** The following table lists the math alphanumeric formatting commands in LaTeX:

<i>Math font</i>	<i>LaTeX command (with example)</i>
italic alphanumeric	<code>\mathit{id}</code>
upright alphanumeric	<code>\mathrm{id}</code>
bold alphanumeric	<code>\mathbf{id}</code>
script alphanumeric	<code>\mathcal{id}</code>
fraktur alphanumeric	<code>\frac{id}</code>
double-struck/open-face alphanumeric	<code>\Bbb{id}</code>
bold symbols	<code>\boldsym{+}</code>

e. MathML; and the verbosity of MathML

MathML is an XML-based mark-up language for mathematical expressions. It is intended to be used in conjunction with (e.g.) XHTML. Since MathML documents are intended to be authored by tools rather than directly (as opposed to LaTeX), a more chatty approach has been taken. However, one does not want to add to this chattiness, which is understandable. For MathML one has looked at the letter-like characters in 10646 (as well as other lists of letterlike items), and now wish to extend upon the number of preformatted letters in 10646. Hence came the “mathematical alphanumeric symbols” proposal for 10646-2. In addition MathML currently defines a large number of “entity names”, i.e. names for characters: `&isrc;` (for script i), `&lscr;` (for script l), `&ifr;` (for fraktur i), etc., etc. Currently most of them refer to code points in the private use zone, but the intent is that the alphabetic ones are to refer to the “math alphanumeric symbols” proposed. MathML also distinguishes between “presentation mark-up” and “content mark-up”. So there are `<mi>`-tags (presentation mark-up identifiers), `<ci>`-tags (content

mark-up identifiers), <mn>-tags (presentation mark-up numerals), <cn>-tags (content mark-up numerals), <mo>-tags (presentation mark-up operators), <co>-tags (content mark-up operators), plus a host of tags to compose these into complex expressions. For upright and bold identifiers, there are also attributes ('fontstyle' and 'fontweight') to control this, *in addition* to the preformatted bold, italic, etc., letters now proposed.

<i>Math font</i>	<i>Current MathML markup (with example)</i>	
italic alphanumeric	<mi>id</mi>	<ci>id</ci>
upright alphanumeric	<mi fontstyle="normal">id</mi>	<ci fontstyle="normal">id</ci>
bold alphanumeric	<mi fontweight="bold">id</mi>	<ci fontweight="bold">id</ci>
script alphanumeric	<mi>&isr;&dscr;</mi>	<ci>&isr;&dscr;</ci>
fraktur alphanumeric	<mi>&ifr;&dfr;</mi>	<ci>&ifr;&dfr;</ci>
double-struck/open-face alphanumeric	<mi>&iopf;&dopf;</mi>	<ci>&iopf;&dopf;</ci>
italic numeric	<mn fontstyle="italic">12</mn>	<cn fontstyle="italic">12</cn>
upright numeric	<mn>12</mn>	<cn>12</cn>
bold numeric	<mn fontweight="bold">12</mn>	<cn fontweight="bold">12</cn>
ordinary symbols	<mo>+</mo>	<co>+</co>
bold symbols	<mo fontweight="bold">+</mo>	<co fontweight="bold">+</co>

f. Suggested future development of MathML

The MathML group should be recommended to make a change to future versions of MathML so that it follows LaTeX's lead in this regard. This way **also MathML systems would have no need whatsoever of any preformatted "math alphanumeric symbols" at separate code points from ordinary alphanumeric characters.** And that without making MathML at all more verbose than it is, on the contrary, it could even be a bit less verbose (see below). In addition the preformatted "math" alphabetic characters already present in 10646/Unicode should not be used in MathML. The new mark-up suggestion below is also unlimited in which letters that may be used in identifiers (if the fonts used actually cover those letters). Note that **in some areas of mathematics it is common to use multi-letter identifiers, often taken from words in a natural language.** In the following table suggested mark-up for next-generation MathML is given (the new tag names are of course up to the MathML community, these are just suggested tag

names; but note the simplification compared to the mess above):

<i>Math font</i>	<i>Suggested new MathML markup (with example)</i>	
italic alphanumeric	<code>&lt;mi&gt;i&lt;/mi&gt;</code>	<code>&lt;ci&gt;i&lt;/ci&gt;</code>
upright alphanumeric	<code>&lt;mr&gt;i&lt;/mr&gt;</code>	<code>&lt;cr&gt;i&lt;/cr&gt;</code>
bold alphanumeric	<code>&lt;mb&gt;i&lt;/mb&gt;</code>	<code>&lt;cb&gt;i&lt;/cb&gt;</code>
script alphanumeric	<code>&lt;ms&gt;i&lt;/ms&gt;</code>	<code>&lt;cs&gt;i&lt;/cs&gt;</code>
fraktur alphanumeric	<code>&lt;mf&gt;i&lt;/mf&gt;</code>	<code>&lt;cf&gt;i&lt;/cf&gt;</code>
double-struck/open-face alphanumeric	<code>&lt;md&gt;i&lt;/md&gt;</code>	<code>&lt;cd&gt;i&lt;/cd&gt;</code>
italic numeric	<code>&lt;mj&gt;12&lt;/mj&gt;</code>	<code>&lt;cj&gt;12&lt;/cj&gt;</code>
upright numeric	<code>&lt;mn&gt;12&lt;/mn&gt;</code>	<code>&lt;cn&gt;12&lt;/cn&gt;</code>
bold numeric	<code>&lt;mm&gt;12&lt;/mm&gt;</code>	<code>&lt;cm&gt;12&lt;/cm&gt;</code>
ordinary symbols	<code>&lt;mo&gt;+&lt;/mo&gt;</code>	<code>&lt;co&gt;+&lt;/co&gt;</code>
bold symbols	<code>&lt;mp&gt;+&lt;/mp&gt;</code>	<code>&lt;cp&gt;+&lt;/cp&gt;</code>

#### g. Search/match

It has been argued that having special characters for “math alphanumeric symbols” would make searching for particularly styled math identifiers easier. Why would it be easier to search for `&ibold;` or whatever (plane 1) code that stands for, than searching for `<mb>i</mb>` or `<cb>i</cb>`? Note that MathML currently allows also for `<mi fontweight="bold">i</mi>` to designate a bold identifier named *i*. Note also that the bold/fraktur/etc. property, according to the proposal in point f, is right next to name of the identifier, not somewhere further out in the surrounding text. We fail to see how the “math alphanumeric” proposal would simplify search at all. It actually makes it more difficult to find all occurrences of a name, since one would also need to consider several kinds of preformatted versions of letters as well. That leads to complications that are unlikely to be satisfactorily solved in most search software. Indeed, one of the reasons that XML is said to improve things is it enables searches for particularly tagged data (like street name in an address, or a fraktur identifier in a math expression). The “math alphanumeric symbol” characters proposal run contrary to what is otherwise claimed for XML documents, including MathML documents.

#### h. “Plain text math”; or necessity of mark-up for math

There are suggestions to have some kind of “plain text math”. The suggestion relies on having new control codes for things like the subscript command (`_`) in LaTeX, the `\over` command, etc. The suggestion requires that the “plain” text be parsed. This is not really plain text, but marked-up text, even though the mark-up consists of control codes. However, this approach does not require the allocation of “math alphanumeric” characters either. Just use a set of control codes that correspond to the LaTeX commands listed above in point d and the suggested mark-up tags in point f above.

j. Misuse

Whatever the limitations set up, the proposed “math alphanumeric symbols” can *and will* be misused to make plain text italic, bold, etc. This will work well only for English (and possibly Greek), due to the limitation of available preformatted letters.

k. Conclusion

The preformatted compatibility characters in 10646 must not be allowed to lead to the acceptance of the “mathematical alphanumeric characters”. There are much better alternatives. The MathML designer community should be given the advice sketched in point f above. **Preformatted letters (digits) have never been needed for math before, and there is no need, nor any advantage, to introduce them now. The existing preformatted letters in 10646 should ideally never have been introduced, and should not be used in any application, math oriented or otherwise. If an identifier (or operator) is in bold, italic, fraktur, etc. is significant in math expressions. However, this does not imply that that kind of distinctions should be made at the character level. There are much better, and more general, ways of dealing with these distinctions as shown above. Please do not do the disservice to the math community of accepting the suggested preformatted letters and digits.** Note also that the suggestion in point f above also covers bold-face operators.

-----end of N2168-----