# "It's Volatile": Standards-Based Research & Research-Based Standards Development

**Walsh, John A.**
jawalsh@indiana.edu
Indiana University

**Hooper, Wally**
Indiana University

---

*You even have
my field guide. It's you I love.
I have believed so long
in the magic of names and poems
I hadn't thought them bodiless
at all. Tall Buttercup. Wild Vetch.
"Often I am permitted to return
to a meadow." It all seemed real to me
last week. Words. You are the body
of my world, root and flower, the
brightness and surprise of birds.
I miss you, love. Tell Leif
you're the names of things.*
—Robert Hass, "Letter"

*It's volatile because anciently painted
with wings in this manner whence came
this character* ☿ *for mercury.*
— Sir Isaac Newton, "Praxis,"
Babson Collection (Burndy Library Collection)
MS. 420, Huntington Library

Digital humanities scholarship often integrates humanities scholarship (literary studies, historical studies, and so on) with technological research and development. Some of this technological work takes the form of standards development. The most noteworthy example of such standards development in the digital humanities community is the Text Encoding Initiative (TEI). The TEI provides Guidelines for encoding texts for scholarly and general use. The TEI is pervasive in digital humanities and digital library contexts. It is a de facto standard developed and evolved over the past twenty some years through the efforts of a number of dedicated scholars, librarians, and technologists, and with input from the larger community of TEI users.

Another standard of significance to the digital humanities community is Unicode. Our paper presents a case-study of a successful effort to have included in the Unicode standard dozens of characters required by the *Chymistry of Isaac Newton*, an ongoing digital humanities project to digitize and edit, study and analyze the alchemical works of Isaac Newton and to develop various scholarly tools around the collection. Unicode has become the universal character encoding standard. Unicode is nothing more, as it is certainly nothing less, than a massive mapping of characters to numbers, a mapping that seeks to accommodate all the world's languages and writing systems, including symbols of all sorts—mathematical symbols and operators, astronomical and astrological symbols, Zapf Dingbats, and many more. Operating systems, and the applications built upon them—databases, word processors and text editors, browsers, graphics software, and games—depend on such mappings, or encodings, to reliably reference, store, input, output, and display textual data. The Unicode Consortium's "What is Unicode" page `http://unicode.org/standard/WhatIsUnicode.html` accurately reports the standard's significance: "Unicode is required by modern standards such as XML, Java, ECMAScript (JavaScript), LDAP, CORBA 3.0, WML, etc., and is the official way to implement ISO/IEC 10646. It is supported in many operating systems, all modern browsers, and many other products. The emergence of the Unicode Standard, and the availability of tools supporting it, are among the most significant recent global software technology trends."

In spite of Unicode's impressive comprehensiveness, it does not include every character ever used. It does not at present, for instance, include many of the alchemical symbols found in Isaac Newton's alchemical writings. Unicode provides a "private use area," a series of reserved *code points* (the numbers assigned to characters) for projects and products to use "privately" for mapping to characters not represented in Unicode. A project like the *Chymistry of Isaac Newton* can make use of this private use area to map to characters that are not already described in the standard. A pitfall of the Private Use Area is that it is meant to be used privately; it is not suitable for easily interchangeable or interoperable data.

One project's implementation of the Private Use Area could conflict with another project's. And fonts would not typically include characters for Private Use Area code points, since by their nature these codepoints are not assigned permanently to any one character but are perpetually open for *private* assignment, not as part of the public standard.

So when a project stumbles upon a rich collection of important characters and symbols that are relevant and useful beyond the interior confines of one's own project, one can make a significant scholarly contribution by documenting and describing these characters and proposing them for inclusion in the Unicode encoding standard. The alchemical symbols so common in Isaac Newton's chymical manuscripts, are common also throughout manuscript and print alchemical literature. The graphically and semantically rich symbols also have potential utility in design, computer art, and even gaming applications. Even the few symbols that are potentially unique to Newton are worthy of consideration in the Unicode standard, given Newton's stature as one of the giants of science and the vast wealth of scientific, historical, biographical, and popular literature related to Newton.



Figure 1. Basil Valentine. "A Table of Chymicall & Philosophicall Charecters with their signs." *The Last Will and Testament of Basil Valentine*, 1671. These and other symbols are commonly found in Newton.

The process by which one moves a Unicode proposal through the development, review, and approval process is formal and rigorous. It is very rewarding in fostering a better understanding of one's source material and in pointing the way to undiscovered or avoided basic research questions. To encode and identify characters and symbols, one must name the things, and naming is indeed a very difficult and powerful task, a task often challenged and enriched by puzzling ambiguity and obscurity. The process is very rewarding also because it is very much peer-reviewed. Our proposal greatly benefited from an iterative review and excellent advice, challenging questions, and constructive criticism from a number of very smart, helpful, interested experts serving on the Unicode Technical Committee (UTC).

Our paper provides a case-study of one project's navigation through the Unicode proposal, review, and approval process. We also provide a more theoretical discussion, illustration, and examination of the mutually beneficial relationship between technical standards development and basic humanities research.

## References

**Unicode Consortium** (15 June 2009). *What is Unicode?*. `http:/ /unicode.org/standard/Wh atIsUnicode.html` (accessed 15 Nov 2009).

**Newman, William R. (ed.)** (9 May 2008). *The Chymistry of Isaac Newton.* `http://www.ch ymistry.org/` (accessed 15 Nov 2009).