

Emerging Technologies Research Paper: Topic Proposal

Elliot Chandler (Wallace)

February 14, 2017

Electronic document representation models are an emerging technology and an area of complex, ongoing research. Development of these models began with the advent of telegraphy and has continued since then, beginning with early systems such as Morse and Baudot code, and moving through teleprinter control systems and more complex models such as PDF, T_EX, Qt, and PostScript. There are four main contexts for which effective representation models are needed: process-internal, process-external, line transmission¹, and user interface. The chief problems with current models (mostly multi-layered model stacks such as HTML+Unicode) are:

- Many overlapping models
- Many domain-specific models, each with their own drawbacks
- Advances in models are not always shared between various application fields (e.g. a legacy structured data format might require use of ASCII,

¹The distinction between a document representation model used in line transmission and a line transmission protocol is somewhat blurry.

EBCDIC, or another obsolete character encoding system, and not inherit advances made in that field such as by Unicode)

- No consistent interchange model for documents (so, e.g., document content stored using a page description language such as PostScript may not transfer losslessly when used in a GUI toolkit such as Qt GUI)

These drawbacks result in duplicated effort, unnecessary parser, renderer, and authoring/editing² complexity, and frustration for users when software does not behave consistently or interoperate well.

My proposed research project is to investigate ways that some or all of these challenges can be ameliorated or resolved, and to study previous and ongoing work in this area.

In my research, I should remember:

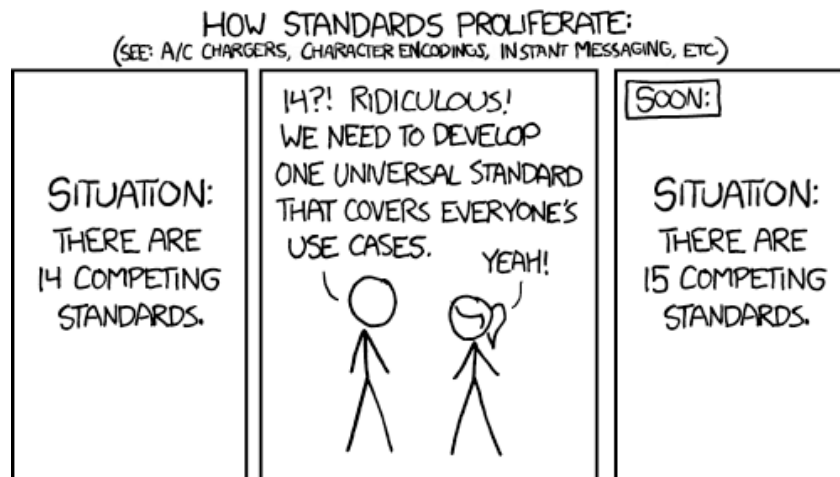


Figure 1: “Standards” (CC-BY-NC 2.5; <https://xkcd.com/927/>)

²(both plain-text authoring/editing such as is typical for source code and markup languages, and GUI editing as is typical for graphics files and word-processing documents)