IEEE Internet Computing

www.computer.org/internet computing

The Good and the Bad of Top-Level Domains

Barry Leiba

Vol. 13, No. 1
January/February, 2009

Editor: Barry Leiba • leiba@watson.ibm.com

The Good and the Bad of Top-Level Domains

Barry Leiba • *IBM Research*

ICANN has recently announced an "open season" on top-level domains, to start some time in 2009. This will dramatically expand the namespace for Internet domain names, and will allow cities, industries, and companies to register specific top-level domains for themselves. What effect will that have on the companies involved, and on the Internet users? In this column, the author explores that question and gives his opinion on the answer.

When the Internet began, there were as many top-level domains as fingers on your hand. That handful comprises the ones we're all still most familiar with today: *com*, *org*, *edu*, *gov*, and *mil* (well, *mil* isn't very well-known, true). But let's back up.

What's in a Name?

A *fully qualified domain name* (FQDN) names a computer, a service, or set thereof, on the Internet. If you read my blog, for instance, you'll see staringatemptypages.blogspot.com in your browser's address field; that's my blog's FQDN. That domain name represents a hierarchy, and the top of the hierarchy — the narrowest part, the focal point — is at the end. In this case, staringatemptypages.blogspot.com has no *subdomains*; it is at the bottom of the hierarchy. There are lots of subdomains of blogspot.com besides staringatemptypages.

Now, blogspot.com is a domain name owned by Google (nowadays), and because they own that, they also own all lower-level domains, all subdomains (including the one for my blog). Because of the way things are assigned, all of the subdomains of blogspot.com are under the same administration. This doesn't have to be the case. A company could register, say, example.com (actually, not example.com, which I'm using here as ... an example; see the discussion later about RFC 2606) and then act as a registry for subdomains, giving administrative control to the sub-registrants, so that barry.example.com and jane.example.com were entirely separate, administratively.

If we peel another layer off, we get to *com*. Because there's nothing above it in the hierarchy, we call it a top-level domain (TLD). The assignment of subdomains to top-level domains (we call such subdomains second-level domains) is a task given to registries. The whole thing is managed by an organization in the US called the Internet Corporation for Assigned Names and Numbers (ICANN), which has assigned the various TLDs to different registry managers.

Anyway, back in the old days, long before ICANN, the TLDs had specific meanings and were assigned based on how the organization requesting it fit the fairly specific definitions:

- *com*: Commercial entities. Note that this didn't equate to commercial use of the Internet, which was strongly discouraged then. Until the early 1990s, "dot com" usually meant a research arm of a company, or some branch that had a joint project going with a university or government. Currently managed by VeriSign.
- *org:* Noncommercial organizations that didn't fit in another category. Currently managed by the Public Interest Registry.
- *edu:* US educational organizations, such as colleges and universities. Currently managed by Educause.
- *gov:* Nonmilitary US government orga-

 1089-7801/09/$25.00 © 2009 IEEE

Glossary

- *Country code TLD:* two-character top-level domain (see below) that is associated with a specific country, using the country's ISO (see) country code (such as us and fr).
- *Domain name:* a hierarchical name assigned to a computer or set of computers on the Internet, such as research.example.com. The most specific level of hierarchy appears first, and the hierarchy levels are separated by periods, spoken as "dot."
- *Domain name registry:* list, maintained by a registry manager, of a certain subset of domain names, usually in a portion of the hierarchy assigned by ICANN (see below) to the registry manager.
- *Domain Name System (DNS):* an Internet service that translates domain names to numeric Internet addresses. The DNS also associates certain other information with domain names.
- *Fully qualified domain name (FQDN):* the domain name (see above) that completely identifies a computer or service, and for which no more specific hierarchy level is needed. A computer called server1 might have an FQDN of server1.research.example.com.
- *Internet Corporation for Assigned Names and Numbers (ICANN):* a nonprofit corporation chartered to manage certain functions for the Internet on behalf of the US government. For the purpose of this discussion, their function is to manage the assignment of domain names and Internet addresses.
- *International Organization for Standardization (ISO):* a standards organization comprising representatives from national standards organizations of the member countries (most of which have strong ties to their respective governments).
- *Link farm:* a Web site that offers little or nothing more than a list of links, often about a particular topic, and often used to generate revenue through a pay-per-click system or to manipulate search-engine results.
- *Parked domain:* an otherwise unused domain name set up to receive service requests and usually to return some sort of advertisement. Registry managers often park domain names to entice people to pay to register them, or to use unregistered domain names to create advertising revenue with link farms (see above). When other parties do this, it might amount to domain tasting or cybersquatting.
- *RFC 2606:* a "best current practices" document that defines certain domain names and top-level domains as "reserved" for purposes such as testing and examples (http://tools.ietf.org/html/rfc2606).
- *Second-level domain:* a domain name that includes one hierarchy level above the top-level domain (see below), such as example.com. Second-level domains often, though not always, identify a separate administrative entity, such as a company, a school, or a government agency.
- *Subdomain:* a domain name that's at a more specific hierarchy level than another. The term is usually, but not always, used when descending one level of hierarchy. So research.example.com is a subdomain of example.com.
- *Top-level domain (TLD):* the final, most general hierarchy level for domain names (such as *com* and *org*).

nizations. Currently managed by the US General Services Administration.
- *mil*: US military organizations. Currently managed by the US Department of Defense Network Information Center.

The TLD *net* was added early on (now managed by VeriSign) for network providers, and the resulting six served us well for quite some time, until 2001 (we'll ignore the country code TLDs here; more about those later).

But as the World Wide Web came about and things exploded, in the early 1990s, domain name registration exploded as well. The TLDs *edu*, *gov*, and mil have remained restricted to their original uses, leaving *com*, *org*, and *net* to cover everything else — and they're all pretty much unrestricted and interchangeable, though org still carries some small noncommercial connotation.

It became common for a new company to try to register a domain name and to find that the reasonable variations were already taken. Besides, said some, there's no really good reason we should be limited to those six TLDs. And so in 2001 ICANN defined a bunch more such as, *name*, and *biz*. At the same time, ICANN gave others to registrars that put specific restrictions on their assignment: *travel*, for example, is only used for domains with travel-related functions (such as travel agents and tourist-information Web sites); *museum* is only for domains associated with museums.

Special Use TLDs

These changes really opened things up, and I believe, with some reservations, that it's good to have the special-use TLDs — and more could be added, making it easier for people to find what they're looking for and harder for "bad guys" to fool us. But all this has a bad side, especially when we look at the unrestricted, general-use TLDs.

First, there's the difficulty in sorting out the different TLDs when you're looking for something. One of the most famous cases is with the second-level domain "whitehouse." The US government site for the president and his staff is whitehouse.gov, consistent with the use of the *gov* TLD. But people aren't used to typing ".gov," and they often get it wrong. And whitehouse.com is a pornographic Web site, which has given many an unwary Web surfer a big

surprise. Also, whitehouse.org is, at this writing, a satirical take-off Web site ("The officious Web site of President George W. Bush"), whitehouse.net is a less-clever joke site poking fun at Mr Bush, whitehouse.biz and whitehouse.info at one time both pointed to the same "parked" domain that was just a commercial link farm ... you get the point, yes?

Second, suppose I should found a new company called the "Frobozz Magic Everything Company," and I wanted to register it as frobozzmagic. I see, as I write this, that it's available in *com*, *net*, *org*, *biz*, and *info*, so which do I pick? Well, *com* is the obvious one, because it's the first thing people think of. But maybe my biggest competitor will see that I did that, and will grab the other four and make them point to his company! Maybe I'd better take all five of them, just in case.

On the other hand, suppose I wanted to register frobozzmagicco instead. I see that someone has already taken that domain in the com TLD. I could register my company as frobozzmagicco.biz (or net, or whatever) but do I want to? Will people keep winding up at frobozzmagicco.com when they're looking for me, and get frustrated and give up (or just buy from him, not from me)? Maybe I should just pick a different domain name instead.

And that's the problem, and that's why the new TLDs aren't as widely used as they might be, why they aren't really solving the problem by expanding the reasonable domain name space, and why opening yet more unrestricted, general-use TLDs will do more harm than good.

Even with the restricted, sponsored TLDs, such as the aforementioned travel and museum, there are issues — although they're easier to resolve and have sometimes been resolved quite elegantly. Suppose you want to find the Web site for the Museum of Modern Art in New York City. Its "real" Web site is moma.org, but if you try moma.museum you'll find that the museum TLD gives you some nice help: it lists all of the subdomains of moma.museum, including nyc.moma.museum and sf.moma.museum. The former is the one we want, while the latter is San Francisco's Museum of Modern Art. This is also a real example of the practice of having separately administered subdomains, as shown above with example.com.

The problem, though, is this: would you think to try it? A critical mass of Web sites must use a particular TLD before users will try going there, and that's a significant bootstrapping problem. Once we get to the point at which people are used to looking for museums in the museum TLD, that TLD will be convenient and helpful. Until then, it won't be much more than a novelty.

So it's not clear what the value is in creating more of these specialized TLDs either — though it's a good idea in principle — unless the applicable industry is really willing to make a strong push for a transition to them.

Country Code TLDs

As you look at the list of original TLDs described above, you can see how US-centric things were. Somewhat filling in the gap, and giving different countries at least some level of control, are the country code TLDs. Each country has a two-letter abbreviation that's assigned by the International Organization for Standardization (ISO). Those were turned into top-level domains, with each country's assignments managed at the country's discretion. We see that in the US in what's become a standard for the Web site for each state: www.state.xx.us, with *xx* replaced by the state's abbreviation.

But outside the US, the country code TLDs are the normal way to make domain assignments. Although a French company can certainly go to VeriSign and get a com domain (such as peugeot.com, which exists), it's more common for them to use the *fr* TLD (such as peugeot.fr). Companies such as eBay and Google that have localized sites around the world use the country codes to distinguish the various local versions (google.fr, google.de, google.es, and so forth — and, yes, google.us will take you to the US version, wherever you travel, even when the local DNS resolves google.com locally).

The country codes are problematic; France2 television is at france2.fr, but france2.com used to be an offensive anti-France web site. The UK has chosen to subdivide the *uk* TLD, so businesses there have domains like bbc.co.uk (the "co," here is analogous to "com"), which causes confusion when people try things like bbc.uk.

We also have some interesting artifacts that aren't "problems," but that are curious. The domain del.icio.us makes clever use of the us TLD, and we can find similar examples. Many countries have little use for their own domains — island countries such as Tuvalu (*tv*), and other small countries like Andorra (*ad*) — and find it fruitful to sell domain names for export. There's widespread use of *cc*, the country code for Cocos Islands, and such usage extends far beyond the Indian Ocean.

Again, would you think to look for americanidol.tv, or would you more likely just try americanidol.com? (The latter is the real one; the former is a link farm.) In any case, we're back to the need for defensive registration. If you want to protect your brand, the more TLDs there are, the more domain names you need to register to make sure you're protected.

Are TLDs Obsolete?

We should also ask whether most people care about TLDs anymore. Increasingly, it's search engines, not raw URLs, that are steering us to the Web sites we're looking for. Users can try to guess the URL for,

say, the Jacob Burns Film Center (an independent-film cinema in my area — it's burnsfilmcenter.org, but someone who calls it "the Jacob Burns" would probably not try that), or they can just put "jacob burns movies" into Google, and have it show up at the top of the hit list, not caring what domain name the cinema chose, or whether it's a *com* or an *org*.

In fact, the ease with which search engines handle this function makes even browser bookmarks (or "favorites") obsolete, at least as entry points to the subject Web sites. It's easier just to do a search than to wade through a large pile of disorganized bookmarks, unless you've been meticulous about organizing them in some sensible way.

The harder problem lies in the laps of the merchants, who want you to buy from them rather than from their competitors. A corollary to using searches to find things is that it becomes important for businesses to do what they can to see that they come up near the top of the search results — few users will scroll down very far, and they're more likely to choose a business from among the first few listed.

Then there are the reserved domains. Many standards specifications, as well as other writings, need to use domain names as examples. In this article, I'm using several real domains for illustrative purposes, and that's what I want to do. But if I were just demonstrating how to use a protocol, I would more likely not want to use someone's real domain name — people can sometimes get touchy about seeing a lot of unwanted traffic, especially when it turns out that when you wrote your example using foo.com, you had no idea that there really is such a domain (it's a site to look up people, an acronym for Four One One).

As a way to avoid this problem, the IETF published Best Current Practice (BCP) 32 in 1999. BCP 32, also known as RFC 2606, [1] defines four reserved top-level domains — *test*, *example*, *invalid*, and *localhost* — and three reserved second-level domains — example.com, example.net, and example.org — that are guaranteed never to be assigned so that researchers and developers can always use them for example and testing purposes without concern.

If you've ever wondered why there have been so few top-level domains, how the names within them get assigned, or where those odd two-letter TLDs come from, I hope this has answered your questions.

But it will all soon change. ICANN decided at the end of June to open up the TLD name space, allowing the registration of essentially any name at the top level. Details are still not final, but the new plan is expected to go into effect this year (it's also approved the use of TLDs in non-Western character sets, to go along with the use of them in the past few years in second-level domain names).

What that will mean is that organizations willing to pay on the order of US $185,000 (for the first round; prices will vary over time and will probably go down eventually) can purchase TLDs for their own brands and use, or to subdivide and sell second-level names to others. Cities are considering signing up (expect *.nyc* and *.london*, for example), and many companies are considering adding custom TLDs to their sometimes long list of domain names.

Now that it's approved, the expansion is inevitable, and unwise. It will add more confusion to the already confused state I've described here. Will Starbucks get .starbucks? Will they use that to replace starbucks.com as their Web brand? If so, what will their main URL be? Surely, http://starbucks would, by itself, be confusing (and would violate *Domain Name System* (DNS) resolution rules). Will they also decide that they have to register starbucks.nyc, starbucks.london, starbucks.chicago, and so on? If someone grabs the TLD .coffee, will the company get starbucks.*coffee* too? What about .espresso, .latte, .cappuccino, ...?

A lot of contention will likely occur for some desirable names — consider *.web* or *.news* — with battles among companies seeking to sell off parcels of those name spaces for a good deal more than they paid for the right to do it. Such fights are not good for the Internet as a whole.

In the end, it will be expensive and confusing, making money for those selling domain names but providing little real value. It will be insidious, in that companies will likely feel the need to register names to protect their brands, even when they understand the lack of value. And more than now, even, people looking for Web sites will just rely on search engines to find them.

Reference

1. D. Eastlake and A. Panitz, *Reserved Top Level DNS Names*, IETF RFC 2606, June 1999; http://tools.ietf.org/html/rfc2606.