Brief history of top-level domains and challenges for information professionals

Enrique Orduña-Malea; Isidro F. Aguillo

Posted on *IweTel* on April 17 2020

Enrique Orduña-Malea
https://orcid.org/0000-0002-1989-8477

Universitat Politècnica de València
Camí de Vera, s/n. 46020 Valencia, España
enorma@upv.es

Isidro F. Aguillo
https://orcid.org/0000-0001-8927-4873

Cybermetrics Lab IPP-CSIC
isidro.aguillo@csic.es

Abstract: Top-level domains, in addition to fulfilling the technical function of enabling access to network resources, allow information professionals in general—and content managers in particular—to work with domain names for branding and search engine optimization (SEO) purposes. Likewise, they facilitate the performance of metric analyses thanks to their hierarchical structure. The massive creation of new TLDs that was started in the early 21st century by the *Internet Corporation for Assigned Names and Numbers* (ICANN) opened the possibility of being able to select (or even register) a wide variety of web domains. However, the sources of information about TLDs as well as their management and usefulness for both the commercial sector (SEO, web content management, and digital analytics) and scholarly community (webometrics) are not well known among information professionals. The objective of this text is to explain the meaning and function of TLDs, to highlight their different categories, and to show their evolution over time, in order to provide useful information for professionals dedicated to the generation, dissemination, storage/retrieval, and analysis of online content.

Keywords: Top-level domains; TLD; Web domains; Internet; Webometrics; Informetrics; ICANN; Information professional; Search engine optimization; SEO.

Resumen: Los dominios de nivel superior (o *top-level domain*, TLD), además de cumplir con una función técnica de acceso a los recursos en red, permiten a los profesionales de la información en general, y de la gestión de contenidos en particular, poder trabajar con el nombre de dominio con fines de branding y SEO. Así mismo, facilitan la realización de análisis métricos gracias a su estructura jerárquica. La creación masiva de nuevos TLDs iniciada a principios de Siglo XXI por la *Internet Corporation for Assigned Names and Numbers* (ICANN) abrió la posibilidad de poder seleccionar (o incluso registrar) una amplia variedad de dominios web. Sin embargo, las fuentes de información sobre TLDs así como su gestión y utilidad tanto para el sector comercial (SEO, gestión de contenidos web, analítica digital) como académico (Cibermetría) no son muy conocidas entre los profesionales de la información. El objetivo de este breve texto es dar a conocer el significado y función del TLD, sus diferentes categorías y su evolución en el tiempo con el fin de proporcionar información de utilidad para los profesionales dedicados a la generación, difusión, almacenamiento/recuperación y análisis de contenidos online.

Palabras clave: Dominios de primer nivel; TLD; Dominios web; Internet; Cibermetría; Informetría; ICANN; Profesional de la información; Optimización en buscadores; SEO.

1

1. Domain Name System (DNS) and Top-Level Domains (TLDs)

Each device (desktop computer, laptop, tablet, mobile phone, etc.) connected to the Internet is assigned a unique identifier, called an *Internet protocol address* (IP address). The purpose of this identifier is twofold: it allows devices to identify each other, and to exchange information via the Internet (**Shimanoff**, 2013). Webpages are the most frequently exchanged type of electronic information.

Although final users could employ (numeric) IP addresses directly within URLs to access resources served by such devices connected to the Internet through a web browser (e.g., http://128.97.27.37 redirects users to http://www.ucla.edu, the *University of California Los Angeles*'s official website), it is uncommon for them to do so. IP addresses are designed for computers, not for humans, who find these numbers difficult to remember. Moreover, IP addresses are inefficient for branding purposes, as they do not transmit information about websites' content.

The translation of IP addresses to more mnemonic names dates back to the early days of the Internet, where a text file (hosts.txt), designed by Elizabeth Feinler (then at the *Network Information Center*, now *Stanford Research Institute International*) and maintained by Jonathan Postel (*University of Southern California*'s *Information Sciences Institute*), mapped host names to the numerical addresses of computers. As the network started to grow, other solutions were needed. Therefore, **Mockapetris** (1983a; 1983b) created the *Domain Name System* (DNS) in 1983, which operates in the application layer of the *Open Systems Interconnection* (OSI) model. The DNS was subsequently reviewed (**Mockapetris** 1987a; 1987b; **Elz**; **Bush**, 1997) and its implemented planned (**Postel**, 1984a; 1984b; 1994; **Postel; Reynolds**, 1984).

The *Domain Name System* (DNS) translates numerical IP addresses (e.g., 128.97.27.37) into domain names (e.g., www.ucla.edu). In this way, domain names not only serve as a mnemonic interface between humans and machines, but also provide a more persistent identifier, as IP addresses might change internally while keeping the domain name invariable for final users (**Mueller**, 1998).

The DNS operates in a tree-like hierarchy to structure the domain name. At the top of the hierarchy is the authoritative "root" computer server. Then, from right to left, one sees the resource ID (top-level domain), the network/server name (second-level domain), and the service name (third-level domain), all of them separated by dots (**Christou**; **Simpson**, 2009; **Smith III**, 2013). In addition to this, the server name can generate subdomains (Fig. 1).
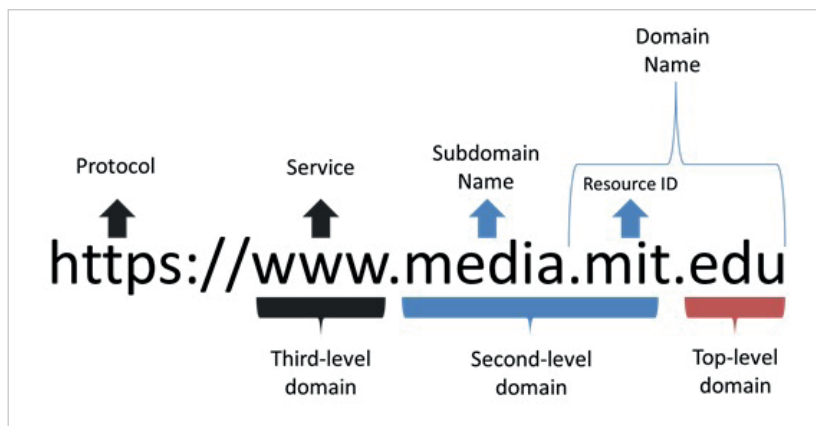

Figure 1. Anatomy of the *Domain Name System*

Domain names are meant to cover from the broadest level of the organization at the right-hand side to the most specific level at the left-hand side (**Halvorson** *et al.*, 2012). Consequently, the DNS can identify the Internet presence of different social entities (e.g., brands, companies, products, individuals, etc.), among which academic entities (universities, journals, scientific organizations, thematic repositories, author personal webpages, etc.) stand out.

Both the hierarchical nature of DNS and the organizational purposes of top-level domains allowed the emergence of new quantitative studies on the Internet in general, and the web in particular, especially when search engines started offering advanced search features. Great efforts were made to measure the whole Internet as well as specific environments, among which the academic space stands out (**Orduña-Malea**; **Aguillo**, 2015).

All tasks related to the administration of registries of Internet protocol identifiers, including top-level domains and IP addresses, are known as *Internet Assigned Numbers Authority* (IANA) functions.

Jonathan Postel originally performed these functions manually. However, as the Internet started to expand globally, under President Bill Clinton's administration, the *US Department of Commerce* initiated a process to establish a new organization to perform the IANA functions, the *Internet Corporation for Assigned Names and Numbers* (ICANN), established in September 1998 (**Lindsay**, 2007).

ICANN was under contract with the *US Department of Commerce* until 2009, and it is currently registered as a private non-profit corporation in California (**Mac-Síthigh**, 2010). ICANN is structured on a multi-stakeholder model, including "registries, registrars, Internet Service Providers (ISPs), intellectual property

advocates, commercial and business interests, non-commercial and non-profit interests, representation from more than 100 governments, and a global array of individual Internet users" (**Smith III**, 2013).

At the time when ICANN started its functions (1998), only seven generic top-level domains (.com, .net, .edu, .gov, .mil, .org, and .int) were available, as well as the geographic top-level domains labeled with the two-letter country codes from ISO-3166 (with the exception of the United Kingdom, which uses .uk instead of .gb). **Postel** (1994) manifested that it was "extremely unlikely that any other TLDs will be created."

However, the addition of new top-level domains to the root system was on the table. ICANN formed the *Domain Name Supporting Organization* (DNSO) advisory body in 1999 to study the issues surrounding the formation of new generic top-level domains (**Farley**, 2007; **Halvorson** *et al.*, 2012), which resulted in the publication of a report which set out a rough consensus regarding the creation of new generic top-level domains.
http://www.dnso.org/dnso/notes/20000321.NCwgc-report.html

The creation of new global top-level domains resulted in arguments both for and against (**Mueller**, 1998; **Halvorson** *et al.*, 2012; **Smith III**, 2013; **Mahler**, 2019). On the one hand, the resolution of name conflict problems, the global competition in Internet-related services, the creation of second-level domain names that were more responsive to user demands, or new opportunities for entities that had been shut out under the previous name structure were received as positive points. However, the possibility of name speculation, fraud (cybersquatting, typosquatting), or conflicts with trademark law were perceived as negative points.

Finally, in 2000, ICANN selected seven top-level domain proposals, representing the first addition of TLDs to the Internet since the 1980s. The selected TLDs were .aero (air-transport industry), .biz (businesses), .coop (cooperatives), .info (all uses), .museum (museums), .name (individuals), and .pro (professions).
https://www.internic.net/faqs/new-tlds.html

In addition, .asia, .cat, .jobs, .mobi, .tel, .travel, .post, and .xxx were additionally included in 2004, bringing the total number of top-level domains to 22.

In 2005, ICANN's *Generic Names Supporting Organization* (GNSO) began a policy development process to consider the introduction of new gTLDs. A draft final report on the introduction of new generic top-level domains was published in 2007, including a set of recommendations.
https://gnso.icann.org/en/drafts/pdp-dec05-draft-fr.htm

ICANN adopted these specific policy recommendations and developed an applicant guidebook describing the program's requirements and evaluation processes.
https://newgtlds.icann.org/en/applicants/agb

ICANN's Board of Directors approved the guideline in 2011, authorizing the launch of the program.

This "New gTLD Program" allowed any public or private entity to apply to register nearly any word, in nearly any language, as a generic top-level domain, provided that the entity could demonstrate the ability to meet certain technical, operational, financial, and other criteria (**Mac-Síthigh**, 2010; **Shimanoff**, 2013). The new environment is shaped by domain name registries (that manage the entire database of all domain names under a specific top-level domain) and domain name registrars (that assign domain names directly to end users).

ICANN began accepting applications for the new top-level domains in January 2012 and released the first set of initial evaluation results to applicants and the public in March 2013. The first new generic top-level domains were delegated in October 2013.

Finally, ICANN unveiled which companies, organizations, start-ups, geographical regions, and others had applied for gTLDs and which domain names they sought during a London news conference on 13 June 2012, known as "Reveal Day."
https://www.icann.org/en/system/files/press-materials/advisory-06jun12-en.pdf

Nearly 2,000 applications for approximately 1,200 different gTLDs were received, including famous brand names, terms denoting industries and professions, terms denoting goods and services, and geographic locations (**Shimanoff**, 2013).
https://newgtlds.icann.org/en

Among the new gTLDs, one can distinguish those related to cultural and linguistic areas. In Spain we find .gal (registered by *Asociación puntoGAL* in 2014 for Galician culture) and .eus (registered by *Puntueus Fundazioa* in 2014 to disseminate Basque Country culture). The case of .cat is different, as this TLD already existed before the launch of the new gTLD (being registered as a sponsored TLD in 2005 by *Fundació PuntCat*). Outside Spain we can find .pyc (registered by *Rusnames Limited* in 2014), .arab (registered by the *League of Arab States* in 2017), .irish (registered by *Binky Moon* in 2014), and .lat (registered by *ECOM-LAC* in 2015). Of interest, other gTLDs were not approved (e.g., .thai) or were withdrawn (e.g., .zulu).

The first new generic top-level domains were delegated by October 2013. Since then, 1,228 new generic-top level domains have been delegated (Figure 2).

With this new environment, a total of 1,584 top-level domains are currently (February 2020) available: 1,240 generic top-level domains (gTLDs; directly managed by ICANN), 315 country-code top-level domains (ccTLDs; managed by each country's domain name regulation corporation), 14 sponsored top-level domains (sTLDs; managed and sponsored by private agencies or organizations that establish and enforce usage rules), 11 test top-level domains (tTLDs), 3 generic restricted top-level domains (grTLDs), and 1 infrastructure top-level domain (.arpa).

Figure 2. Number of new generic top-level domains delegated, broken down by year.
Source: ICANN (*https://newgtlds.icann.org/en/program-status/statistics*)

As a special case, one can distinguish the internationalized top-level domains (IDN TLDs), which contain characters outside of the 37-character letters, digits, and hyphen (LDH) range, including non-Latin character sets (e.g., Arabic, Cyrillic, Hebrew, Chinese, Japanese, Korean, etc.). Currently there are 166 IDNs.

2. Challenges for information professionals

From an information professional point of view, the current diversity of top-level domains offers opportunities to content managers to develop tailored strategies for search engine optimization (SEO), search engine marketing (SEM), and content discovery using customized TLDs, especially for large international companies or organizations.

Appropriate selection of a domain name might help to increase user attention, to cleverly distinguish one website from other similar competing websites, to obtain clicks and visits, and sponsored searches. However, for this tactic to be effective, users should perceive the selected TLD as reputable and memorable. Otherwise, search engine algorithms might treat TLDs differently, depending on how much spam or clicks these TLDs tend to host or receive, respectively.

While Internet freedom allowed the registration of any desired TLD (regardless of the intended purpose of this domain), this freedom has resulted in a chaotic situation from an informational point of view.

SEO companies used to recommend .com, as it is well known by users, for example. However, the use of overexploited TLDs in some sectors might result in negative experiences, especially when they are used for purposes different from what it is expected (for example, .com is expected for commercial companies, not personal websites). The geographic TLDs can also introduce some degree of uncertainty (for example, what type of website is expected for .es or .fr TLDs?). In this sense, the new generic TLDs can add not only customization and branding names for websites, but also some order and information to search engines and analysis.

Precisely, from a webometric point of view, the current diversity of top-level domains offers opportunities for the development of more-specific field research beyond the current filtering system based on geographical or institutional domains. The web is by far the largest hypertextual corpus available for data mining analysis, but its content is huge and lacks proper structure and organization. Efforts to provide metadata, such as *Wikidata*, are not yet sufficient for large data mining projects involving collections of websites with heterogeneous content. To focus on a group of interest, the use of highly specific top-level domains offers important advantages for designing and using robots, crawlers, or other web-scraping tools, as they can be customized to work only inside websites that share one of the specific TLDs.
https://www.wikidata.org

The large size of traditional domains is also a problem for Internet archiving projects beyond the effort done by *The Internet Archive*, which is not flexible enough for research purposes, especially metrics-oriented ones. There are a few national archiving projects linked to country top-level domains (**Brügger**, 2018), but new domains are still to be explored in this regard.
https://archive.org

Artificial intelligence (AI) will certainly be used in the future, not only to extract huge web data content but also to tag it automatically with metadata. However, in the meantime, filtering webspaces by top-level domain is a feasible option.
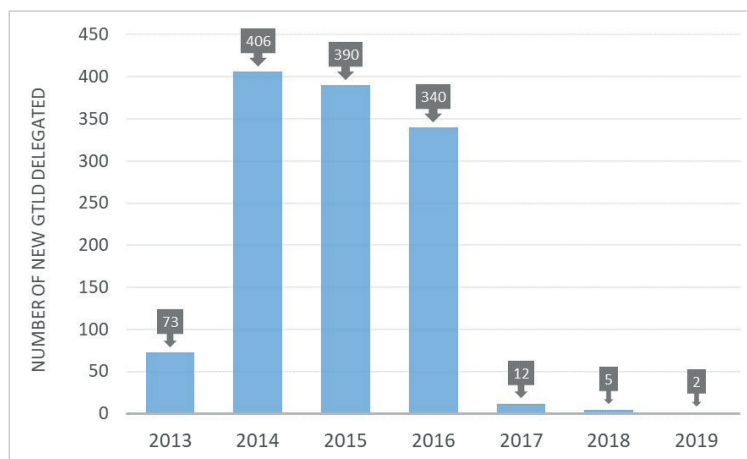
4

3. References

Brügger, Niels (2018). *The archived web: Doing history in the digital age*. Massachusetts: MIT Press. ISBN: 978 0 262 03902 4

Christou, George; **Simpson, Seamus** (2009). "New governance, the internet, and country code top☐level domains in Europe". *Governance*, v. 22, n. 4, pp. 599-624.
https://doi.org/10.1111/j.1468-0491.2009.01455.x

Elz, Robert; **Bush, Randy** (1997). *Clarifications to the DNS specification. Internet engineering task force* (*IETF*). *Network Working Group*. RFC 2181.
https://tools.ietf.org/html/rfc2181

Farley, Christine H. (2007). "Convergence and incongruence: trademark law and ICANN's introduction of new generic top-level domains". *25 John Marshall journal of computer & information law*, v. 625.
https://ssrn.com/abstract=1400304

Halvorson, Tristan; **Szurdi, Janos**; **Maier, Gregor**; **Felegyhazi, Mark**; **Kreibich, Christian**; **Weaver, Nicholas**; **Levchenko, Kirill**; **Paxson, Vern** (2012). "The BIZ top-level domain: ten years later". In: *International Conference on Passive and Active Network Measurement*, pp. 221-230. ISBN: 978 3 642 28537 0
https://doi.org/10.1007/978-3-642-28537-0_22

Lindsay, David (2007). *International domain name law: ICANN and the UDRP*. Portland, Oregon: Hart Publishing. ISBN: 978 1 84113 584 7.

Mac-Síthigh, Daithí (2010). "More than words: the introduction of internationalised domain names and the reform of generic top-level domains at ICANN". *International journal of law and information technology*, v. 18, n. 3, pp. 274-300.
https://doi.org/10.1093/ijlit/eaq007

Mahler, Tobias (2019). *Generic top-level domains: A study of transnational private regulation*. Cheltenham: Edward Elgar Publishing. ISBN: 978 1 78643 514 9

Mockapetris, Paul (1983a). Domain names - concepts and facilities. *Task Force* (*IETF*). *Network Working Group*. RFC 882.
https://tools.ietf.org/html/rfc882

Mockapetris, Paul (1983b). Domain names – implementation and specification. *Task Force* (*IETF*). *Network Working Group*. RFC 883.
https://tools.ietf.org/html/rfc883

Mockapetris, Paul (1987a). Domain names - concepts and facilities. *Task Force* (*IETF*). *Network Working Group*. RFC 1034.
https://tools.ietf.org/pdf/rfc1034

Mockapetris, Paul (1987b). Domain names – implementation and specification. *Task Force* (*IETF*). *Network Working Group*. RFC 1035.
https://tools.ietf.org/pdf/rfc1035

Mueller, Milton L. (1998). "The battle over Internet domain names: Global or national TLDs?". *Telecommunications policy*, v. 22, n. 2, pp. 89-107.
https://doi.org/10.1016/s0308-5961(97)00062-1

Orduña-Malea, Enrique; **Aguillo, Isidro F.** (2015). *Cibermetría. Midiendo el espacio red*. Barcelona: UOC Publishing. ISBN: 978 84 9064 654 0.

Postel, Jonathan B. (1984a). Domain name system implementation schedule - Revised. *Network Working Group. Request for Comments: 921*.
https://tools.ietf.org/html/rfc921

Postel, Jonathan B. (1984b). *Domain name system implementation schedule. Network Working Group. Request for Comments: 897*.
https://tools.ietf.org/html/rfc897

Postel, Jonathan B. (1994). *Domain name system structure and delegation. Network Working Group. Request for Comments: 1591*.
https://tools.ietf.org/pdf/rfc1591

Postel, Jonathan B.; **Reynolds, Joyce K.** (1984). Domain requirements. *Network Working Group. Request for Comments: 920*.
https://tools.ietf.org/html/rfc920

Shimanoff, Eric J. (2013). "The dot times they are a-changin': how new generic top level domains (gTLDS) will change consumer perception about the internet". *Cardozo arts &* entertainment, v. 32, pp. 891-926.
https://www.cll.com/media/publication/27_Shimanoff%20Article.pdf

Smith III, Joseph P. (2013). "The tangled web: A case against new generic top-level domains". *Richmond journal of law & technology*, v. 20, n. 3, pp. 1-42.
https://jolt.richmond.edu/2014/06/23/the-tangled-web-a-case-against-new-generic-top-level-domains