

Please note: This is the accepted version of the article published online before print, April 28 2016, New Media & Society, doi: 10.1177/1461444816643790

What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain

Anat Ben-David

The Open University of Israel

Abstract

This article argues that the use of the Web as a primary source for studying the history of nations is conditioned by the structural ties between sovereignty and the Internet protocol, and by a temporal proximity between live and archived websites. The argument is illustrated by an empirical reconstruction of the history of the top-level domain of Yugoslavia (.yu), which was deleted from the Internet in 2010. The archival discovery method used four lists of historical .yu URLs that were captured from the live Web before the domain was deleted, and an automated hyperlink discovery script that retrieved their snapshots from the Internet Archive and reconstructed their immediate hyperlinked environment in a network. Although a considerable portion of the historical .yu domain was found on the Internet Archive, the reconstructed space was predominantly Serbian.

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

Keywords

Web history, Web archives, Internet Archive, Wayback Machine, Yugoslavia, Serbia, Digital Heritage, ICANN, National Webs, ccTLD

Corresponding Author: Anat Ben-David, Department of Sociology, Political Science and Communication, The Open University of Israel, 1 University Road, P.O.Box 808, Ra'anana 43537, Israel. Email: anatbd@openu.ac.il

Introduction

On 30 March 2010, the country code top-level domain (ccTLD) of Yugoslavia—.yu—was deleted from the Internet's Domain Name System (DNS). Formally, the .yu ccTLD had lost its legitimacy, since the country to which it was delegated in 1989, the Socialist Federal Republic of Yugoslavia (SFRY), no longer existed. By 2008, ICANN (Internet Corporation for Assigned Names and Numbers) delegated new ccTLDs to the former Yugoslav republics that were recognized by the UN as sovereign states, and because Serbia and Montenegro were the last to split, their websites officially operated under the .yu domain until they were delegated the .rs and .me domains.

In 2008, ICANN and the Serbian National Registry of Internet Domain Names announced a two-year transitional phase to allow webmasters enough time to register their websites under the new Serbian domain before the .yu domain was removed

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

(IANA, 2010). Nikola Smolenski, a Web developer and a Wikimedia activist, realized the fragility of the transitional stage. On 17 February 2009, he posted on Wikimedia's Wikibots-L mailing list asking fellow Wikipedians to help him replace all references to .yu URLs in the various pages of the Wikimedia project. The risk, wrote Smolenski, was 'that readers of Wikimedia projects will not be able to access information that is now available to them', and that 'with massive link loss, a large number of references could no longer be evaluated by the readers and editors' (Smolenski, 2009: np). He used a Python script to generate a list of 46102 URLs in the .yu domain that were linked from Wikimedia projects and that had to be replaced. A day before the removal of the domain, he also systematically queried Google for all URLs in the .yu domain per sub-domain, which yielded several thousand results.¹ Smolenski's lists are a last snapshot of the presence of the Yugoslav domain on the live Web. The day after he conducted the search, the .yu ccTLD was no longer part of the Internet root, resulting in the link loss he had anticipated.

Despite the historical significance of the former Yugoslav domain in terms both of social history and of Internet history, four years after the removal of the .yu domain there is little evidence on the live Web that it ever existed. While in many countries the websites hosted in the local domain are archived by national libraries in order to preserve the country's national digital heritage, the disintegration of Yugoslavia during the implementation of its national Web has left its digital heritage unpreserved.

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

Although the Internet Archive contains snapshots of past .yu URLs archived in the years the domain existed, these cannot be retrieved unless the URL addresses are known.

The deletion of the .yu domain serves as an extreme case for Web historiography. Since the live Web no longer contains evidence of its Yugoslav past, the attempt to study its history brings to the fore questions about the appropriateness of the use of the Web, and of the archived Web in particular, as a primary source for historical research. In this paper I bring together Web historiography and national Web studies to elaborate on these questions. By reconstructing the history of the .yu domain from the Internet Archive, I move beyond the discussion about the appropriateness of the archiving process or the selection method to argue that the mere possibility of remembering a national Web's past is constituted by two structural dependencies on the present: first, the structural ties between nation-states and Internet protocol, and second, the dependence of the Web archive on the live Web.

The first dependence involves the DNS protocol. Interestingly, the DNS itself is a memory device, since it translates IP numbers into mnemonic addresses (Mueller, 2004). As a hierarchical and universal system for the resolution of Web addresses, the DNS is the Internet's most strict authenticator of sources: HTTP requests of Web addresses incompatible with the DNS will not resolve. At the same time, the DNS is also the Internet's most strict authenticator of nation-states. The DNS is managed by

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

ICANN, which delegates ccTLDs to countries enlisted in ISO-3166-1, the list of the official names of countries and territories recognized by the UN, and their two-letter suffix (Mueller, 2004). As new countries are added to the list, their newly delegated ccTLDs are added to the DNS and subsequently emerge on the live Web. But when countries dissolve, a removal of a ccTLD from the DNS consequently deletes the possibility of resolving its historical addresses on the live Web. The other side of the protocol of mnemonics is thus permanent memory loss.

The second dependence involves access interfaces to Web archives in general, and the Internet Archive's Wayback Machine (IAWM) in particular.² The IAWM allows for retrieving past snapshots of a single URL and assumes a temporal proximity between live and archived websites (Ben-David and Huurdeman, 2014). As Rogers (2013) notes, the IAWM's premise is that the user first encounters a (broken) URL on the live Web and then consults the IAWM to view its past snapshots. Put differently, one has to *know* the URL of the live Web in order to retrieve its archived snapshot. However as the Web grows older and historical URLs are deleted from the live Web, the possibility of knowing their past is diminished since it is no longer possible to rely on the Web of the present as a starting point to the archive. The ties between the Web archive and the live Web are also evident during users' engagement with the archived websites, as researchers often consult the live Web to validate the content or ownership

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

of an archived URL. The question remains whether archived websites are appropriate primary sources for Web historiography when they are detached from the live Web.

Arguably, the dependence of the live Web on the DNS, and consequently of the Web archive on the live Web, inscribes sovereignty and stability into Web archives and national Web history. Sovereign countries whose historical ccTLDs have expanded over the years enjoy the benefit of the enduring proximity between the live Web and its archiving. At the same time, such inscription of sovereignty jeopardizes the Web histories of unstable domains or non-sovereign states and peoples, whose digital pasts are dotted with rupture and deletion.

I begin by outlining the theoretical foundations of this study, drawing both from Web archiving theory and from methods for the study of national Webs. Next, I outline the formal history of the .yu ccTLD and the transformations in its legal status and ownership, and follow this with an empirical reconstruction of the history of the .yu domain from the Internet Archive. I conclude by discussing the potential and limits of studying the Web's deleted past, and the Web's appropriateness as a primary source for telling its own history.

The Archived Web as a Primary Source and the History of National Webs

The history of the World Wide Web can be told from the perspective of its contestations, one of which revolves around questions of its appropriateness as a source of knowledge (Rogers, 2002). While widely considered as the record of humanity in the

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

21st century, critiques and theorists have warned against the Web's elusive character. Compared to previous authoritative repositories of knowledge—such as libraries, museums and archives—the Web's organizing principles are not hierarchical; rather, they are automated, uncontrollable, and often unknown. The authorship and reliability of sources are doubtful, and the Internet's perpetual change both in content and structure undermines the notion of permanence (Chun, 2008, 2011; Ernst, 2012).

Web archives put forward a specific solution to this impermanence and ephemerality by capturing or 'freezing' websites. In this way, they document a moment in a website's lifetime, aiming at creating stable and enduring containers of the Web's past. National Web archives, in particular, are designed to preserve a country's national digital heritage. Decisions are made about the scope of archiving depending on the archiving institution's policy and the country's legislative environment (Phillips, 2005). Whether by running a broad crawl of the national domain several times a year or by archiving a selection of websites around specific themes and issues, most national Web archives are designed to delineate the portions of the Web that they consider relevant to national history and memory. The Web history of countries without an official national Web archive can be retrieved from the Internet Archive, although studies show national differences in the archival coverage of the Web (Thelwall and Vaughan, 2004).

Thus, while archived websites have become the primary source for Web historiography, their appropriateness is debated. One primary concern relates to

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

ontological differences between archived versions of the live website (Brügger, 2010; Masanès, 2005; Rogers 2013). As Web archiving theorist Niels Brügger (2013) notes, compared to born-digital sources (such as live websites or digitized versions of material that previously appeared in analogue format), Web archives are unique, ‘re-born’, digital objects that face the challenge of being constantly reactive to real-time Web and software changes. In this respect, Web historians should keep in mind that Web archives are always incomplete compared to their online originals, which are unstable and constantly changing. Another matter of concern relates to the unit of archiving and analysis (Brügger, 2009; Schneider and Foot, 2005). The organization of Web archives as collections of single websites has been critiqued for not taking into account certain elements of the live Web, such as its networked environment and dynamic content (Brügger, 2012; Rogers, 2013). Finally, since Web archives are accumulative and often contain different snapshots of the same website, the evaluation of the selection of the appropriate source highly depends on contextual information.

The theoretical concerns about the use of Web archives for historical research are accompanied by methodological challenges. Access to most Web archives is provided through the IAWM’s browsing interface, which allows viewing archived snapshots of a single URL. Since most Web archives are not searchable, researchers face difficulties in performing large-scale analyses of broad social and political phenomena that require more elaborate tools than viewing or browsing through archived

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

Web pages (Ben-David and Huurdeman, 2014; Brügger, 2013; Hockx-Yu, 2014; Rogers, 2013).

Despite the theoretical and methodological challenges, historical network analysis has emerged as a method for studying the Web's past with Web archives (Brügger, 2012). The method, initially developed in the early work of Schneider and Foot (2005) on Web sphere analysis, involves a dynamic selection and archiving of a set of Web pages around a theme or an event, which are subsequently analyzed using a triangulation of hyperlink, content, and qualitative analyses. Historical network analysis thus borrows the practice of hyperlink mapping in Internet research but adds to it the temporal aspect of the changing structure of the network over time. To address the problem of the single-site approach of the IAWM, researchers from the Digital Methods Initiative at the University of Amsterdam have developed tools that repurpose the IAWM's single URL browsing interface into a research tool for studying the historiography of a single website and for a reconstruction of historical networks between interlinked archived pages (Rogers, 2013; Weltevrede and Helmond, 2012). More recently, analysis of archived Web data has expanded its scale from a single website or a selection of interlinked websites to the study of the complete national Web archive as a unit of analysis. For example, Hale et al. (2014) used a snapshot of the .uk domain archived by the Internet Archive between 1996 and 2010 to map the structural evolution of the .uk domain and to characterize historical linking practices of British

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

universities. Similarly, the entire national Web archive of Denmark has been analyzed to study the history of the Danish Web, using metrics such as the size of websites, their geo-location, and hyperlinked structure (Brügger 2015, in press). Researchers also analyzed the Dutch Web archive to assess the extent to which evidence of unarchived pages can be retrieved (Hurdeman et al., 2015).

Work on the national domain as object of study dates back to the early days of Internet research, where the primary focus was the tension between the Web's global and national aspects (Halavais, 2000). Previous studies employed quantitative approaches for characterizing national domain names based on content, links, and technologies (Baeza-Yates et al., 2007; Gomes and Silva, 2005). Network mapping has also been employed to demarcate diaspora networks (including former Yugoslavia) (Diminescu, 2012; Mazzucchelli, 2012).

The following reconstruction of the .yu domain from the Internet Archive thus builds on previous research on Web historiography and national Web studies. It applies historical network analysis and methods for the demarcation and characterization of national domains to evaluate the historical implications of the deleted domain and to evaluate the archived Web as a primary source for historical research.

An archival reconstruction of a deleted domain

The history of the Internet has seen several deletions and re-delegations of ccTLDs of countries in transition (ICANN, 2006). In 1994, ICANN retired the ccTLD

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

of Czechoslovakia, .cs, after it delegated new ccTLDs to the Czech Republic (.cz) and Slovakia (.sk). In 1997, ICANN delegated the .tp domain to East Timor—then under Indonesian control and known under its colonial name Timor Português. Domain registry continued until 2005, after the UN recognized the independence of Timor-Leste, and after ICANN re-delegated .tp as .tl (Maguire, 2003). Similarly, in 2001, the former domain of Zaire, .zr, was re-delegated as .cd after the country had changed its name to the Democratic Republic of Congo (IANA, 2001). Apart from secondary sources that describe these domain transitions, there is little evidence on the live Web of their past existence.

The removal of the Yugoslav Internet is not different from its precedents, apart from the fact that it continued to operate on the Web for over than fifteen years after the country it was originally delegated to, SFRY, had ceased to exist. In many ways, the history of the .yu domain is intertwined with the history of the disintegration of SFRY and the decade of the Yugoslav wars. Not only did the Yugoslav Internet play a significant role in creating a virtual landscape of the memory of the country that dissolved and connecting communities that dispersed (Mazzucchelli, 2012, Pogačar, 2010), it was also precedent in the history of the Web when the Kosovo War was infamously dubbed the First Internet War, in which a conflict was reported online and in real time (Keenan, 2001).

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

While acknowledging the social and cultural significance of the history of the Yugoslav Internet, the reconstruction of the .yu domain presented below proposes a spatial and structural angle to its historiography. Thus, while the social and cultural aspects of the .yu domain are beyond the scope of this paper, and while this study focuses on the evolution of the national contours of the domain, the following narrates the formal history of the .yu ccTLD and the struggles over its maintenance and ownership, which is necessary to contextualize the findings of the historical network analysis below.

Yugoslavia joined the Internet in 1989, after ICANN delegated the .yu domain to the University in Maribor and the Josef Stefan Institute in Ljubljana, Slovenia—then still part of SFRY. In 1992, the UN’s sanctions against Yugoslavia led to the exclusion of the country from the academic network. Following the break-up of SFRY in 1992, ICANN delegated new ccTLDs to Slovenia (.si, 1993), Croatia (.hr, 1993), Macedonia (.mk, 1993), and Bosnia-Herzegovina (.ba, 1996).³ In 1993, Serbia and Montenegro formed the Federal Republic of Yugoslavia (FRY). Following the secession of Slovenia and the delegation of its top-level domain, Ms Mirjana Tasić from the Belgrade Faculty of Natural and Mathematical Sciences requested her colleagues in Ljubljana to transfer the jurisdiction of the .yu domain. After months of dispute, which also involved ICANN and European Internet Services institutions, in 1994 the domain was re-assigned to

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

Belgrade as its rightful owner, since Yugoslavia was still a UN member country (Manojlović, 2014).

Between 1994 and 2000—years marked by wars, sanctions, and regional instability—the University of Belgrade maintained the .yu domain on a voluntary basis, and domain registration was free of charge. However due to the lack of financial support from the state and in light of the growing demand for registering addresses, domain registration had been limited to legal entities, with just one website permitted per institution (Manojlović, 2014).

Reform discussions of the maintenance of the .yu domain began in 2003, after Serbia and Montenegro officially formed a Union State in 2003 (Wass, 2003). However the union lasted only three years, and after the split between Serbia and Montenegro in 2006, ICANN decided to delegate the rightful ownership of the .yu domain to the Serbian National Internet Domain Registry (RNIDS) and the separate .me and .rs ccTLDs to Montenegro and Serbia, respectively (Gakovic and Szymczyk, 2007). On 30 March 2010, ICANN removed the .yu domain from the Internet's root zone (IANA, 2010).

This short review of the formal history of the .yu ccTLD highlights the tension between ICANN's domain policy, national interest, and the challenges that the geopolitical turmoil posed on the regulation of the national domain space of a country in transition. The outcome of the regulatory process that led to switching off the .yu

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

domain is that the live Web does not disclose evidence of its deleted Yugoslav past. However snapshots of websites once hosted in the .yu domain can be found in the Internet Archive, as they have been crawled and captured in real time. In a sense, the Internet Archive compensates for the Internet's DNS amnesia.

However, reconstructing the deleted memory of the entire national domain from the Internet Archive can be a daunting task. The methodological challenge is twofold: first, historical addresses must be found through external sources in order to access the Internet Archive; second, the archival retrieval method has to scale from a single URL to a national domain. The method developed to reconstruct the history of the .yu domain from the Internet Archive deals with both challenges, as outlined in the following section.

Method

The method for reconstructing the .yu domain's deleted past from the Internet Archive combines existing methods in historical network analysis, characterization of national Web domains, and a comparative source critique. It assumes an end-user approach to the IAWM, and therefore accesses the Internet Archive using as starting points four lists of historical URLs in the .yu domain, to which I refer as 'seeds'. Three of the lists were captured from the live Web several weeks before the domain expired by Nikola Smolenski, who, as mentioned above, scraped the .yu URLs referenced from Wikimedia projects and from Google search results of URLs in the .yu domain, and also provided a

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

list of websites registered in the Serbian EUNET domain registry. The source of the fourth list is the former Yugoslav (now Serbian) computer magazine *PC Press*. Since 1996, each December issue of the magazine contained an editorial summary of the year's top 50 Yugoslav websites. The magazine's recommended websites between 1996 and 2010 were retrieved from the IAWM. After aggregating page-level URLs to hosts and removing duplicates within each list, the four lists contained 4810 unique websites that once existed in the .yu domain (see Table 1). Subsequently, each list was characterized based on its composition of sub-domains and on the extent to which it contained unique websites that were not found in other lists. Finally, the archival coverage of the seed lists was evaluated using a custom Python script that automatically accessed the IAWM to examine whether there are archived snapshots at the closest date to July 1st of each year between 1997 and 2010.⁴ The HTML pages of the found snapshots were automatically saved in a designated server, and a separate file logged errors encountered by the script (for example, when a snapshot could not be retrieved due to robots.txt or when the requested URL was not found in the archive). The error log file was used to compare the archival coverage of each of the lists, per sub-domain, across the studied years.

After evaluating the seed lists as the primary sources, a hyperlink discovery method was used to find more .yu pages in the Internet Archive and to reconstruct the evolution of the .yu domain's networked structure over time. The hyperlink discovery

Please note: This is the accepted version of the article published online before print, April 28 2016, New Media & Society, doi: 10.1177/1461444816643790

method was performed using five iterations. The first iteration is the procedure described above for the seed list (distance = 0). After the HTML pages of the found snapshots were saved in the server, the script fetched all outlinks found in them. The .yu URLs that were found among the outlinks were added to the seed list after deduplication. Using the same Python script as in the first iteration, the second iteration then used the expanded seed list from the first iteration (distance = 1). Third, fourth, and fifth iterations repeated until no more links in the .yu domain were discovered. The error log file was used to analyze the archival coverage of the discovered dataset. Subsequently, dynamic network graphs were generated with Gephi to visualize the evolution of the ties between the discovered URLs over time. To evaluate national and international characteristics of the reconstructed domain, additional analyses examined the domain distribution of the found outlinks. Finally, patterns of possible migration of websites from .yu to the Serbian and Montenegrin domains .rs and .me during the transitional phase between 2008 and 2010 were detected, using an automated procedure that searched for identical URLs in the .yu, .rs, and .me domains and analyzed the distance between the year of the last archived snapshot of the URL in the .yu domain and the first archived snapshot of the URL in the .rs and .me domains. Two blind coders evaluated the heuristic by manually viewing the archived snapshots of a sample of 100 random URLs (inter-rater agreement: 87%, Cohen's Kappa: 0.66).

Findings

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

The archival discovery method yielded 17,460 unique websites in the .yu domain, and reconstructed their immediate hyperlinked environment in a network of 150,777 pages (see Figure 1). It is estimated that at its peak, 32,772 domains were registered under the .yu domain (IANA, 2010). While not all registered domains were active, the hyperlink discovery method was able to reconstruct only 53.2% of the deleted domain.

<<**Figure 1 about here**>>

The ability to reconstruct the deleted domain greatly depends on the source lists that were used as starting points. A comparative analysis of the lists as a primary source for Web historiography may be seen as a simulation of the future Web historian's work. Assuming that as time goes by there will be more discrepancy between the live website and its archived snapshots, Web historians may need to contextualize archived Web materials with such born-digital lists that represent different organizing principles and hierarchies and that consequently generate different Web historiographies and networks.

The four lists I used as primary source in this study differ greatly in terms both of shared websites and of diversity of sub-domains (see Table 2). While in the lists obtained from Google and EUNET the commercial sub-domain is dominant, Wikipedia is the most diverse, contributing the majority of organizational, educational and Montenegrin websites hosted under the subdomain cg.yu. Interestingly, the list compiled from *PC Press* magazine contributes unique websites that are not hosted in the .yu domain, but are known to have played a significant role in the Yugoslav Web

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

sphere during the wars, such as *nostalgija.com*, a popular repository of ex-Yugoslav music, or *b92.net*, the website of the internationally funded radio station that was shut down by the Serbian government in 1999, but continued to serve online news from the ground (see Table 3). Due to the temporal proximity of its compilation, the editorial list is thus better able to narrate the early historiography of the Yugoslav Web than the other lists that were captured in 2010, since it includes relevant national websites hosted in generic domains.

<< **Table 2 about here**>>

<<**Table 3 about here**>>

An examination of the archival coverage of the unique hosts in each list further confirms the advantage of Wikipedia as a primary source (see Figure 2). The archival coverage of unique websites contributed by Wikipedia spans over longer periods of time compared to the other lists. Interestingly, the distribution of the lists' archival coverage over time shows that Yugoslav websites captured between 2000 and 2005 have endured longer, which corresponds with the domain's recovery after the Milosevic era and before the final split between Serbia and Montenegro, a point to which I return.

<<**Figure 2 about here**>>

Zooming out, the examination of the archival coverage of the entire discovered dataset shows a dramatic increase from 1997 to 2008, followed by a sharp decrease in 2009 and 2010 (see Figure 3). As previously noted, the coverage is derived from the

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

Internet Archive's response code to a retrieval request of URLs. Since three out of the four initial lists were captured in 2010, the skewed archival coverage demonstrates the dependence of the archive on the freshness of its sources. Put differently, the closer the temporal distance between the archiving date and the archiving snapshot, the better its archival coverage. However, the sharp decrease in coverage from 2008 may be an indication of the transition towards the replacement of the .yu domain with .rs and .me. Indeed, my analysis shows that between 2008 and 2010, about 38% (5583) of the discovered .yu websites have registered parallel domains in the Serbian .rs (88%, 5216) and in the Montenegrin .me (5%, 316). Of the migrated websites, 7% (352) have registered more than one website either in both the Serbian and Montenegrin domains or in combination with one of the domains of Bosnia, Croatia, Macedonia, and Slovenia.⁵

<<Figure 3 about here>>

The evolution of the hyperlinked structure of the data follows a similar pattern. As can be seen in Figure 4, the changes in the density of the links between websites in the .yu domain clearly resonate the decades of wars, sanctions, political instability, and struggles over the ownership and maintenance of the domain described earlier in this article. The stagnation of the .yu domain during the Milosevic era and the Kosovo War is evident in the low average degree (the number of links between each host) of the reconstructed network between 1997 and 1999 (0.3), where the majority of websites were not linked to the national domain. Significant internal linking among .yu websites

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

became apparent only after the end of Milosevic's regime in 2000, rising to an average degree of 1.7 in 2005. The split between Serbia and Montenegro in 2006 is evident in a mild decrease in the average degree of the network in 2007 (1.316), followed shortly after by the dilution of the network in preparation for the replacement of the .yu domain with the new ccTLDs .rs and .me (an average degree of 0.2 in 2010). In other words, both the archival coverage and the internal linking patterns of the .yu domain are closely tied with stability and sovereignty.

The ties between sovereignty and the history of the .yu domain are also evident in linking patterns to external domains. The discovered dataset contains more hyperlinks from .yu to generic domains than to other ccTLDs (see Figure 5). Countries that received the most outlinks from .yu websites host large communities of Serbian immigrants, such as Germany, the UK, and Italy (Bilbija, 2013). Outlinks from .yu to ccTLDs of other former Yugoslav republics are far less frequent, thereby reaffirming the emergence of .yu as a Serbian domain.

<<Figure 4 about here>>

<<Figure 5 about here>>

Discussion

In assessing the ability to reconstruct any domain's past, one should keep in mind that the use of the (archived) Web as a primary source can only date back to 1996, the year of the establishment of the Internet Archive. In that sense, the Web does not remember

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

the history of ccTLDs deleted prior to 1996, such as the Web of Czechoslovakia. It also does not remember the early history of the Web of SFRY, and the archival reconstruction of the Yugoslav domain presented in this study begins after the signing of the Dayton agreement and after the domain had already been transferred from Ljubljana to Belgrade.

The method applied in this study has been able to partially reconstruct the networked history of the .yu ccTLD from the Internet Archive. The partiality of the reconstruction does not necessarily indicate that the undiscovered portions of the domain were not archived. Rather, it may be an indication of the limit of the hyperlink discovery method that is based on the initial seed lists. The completeness of the reconstruction effort could have been aided by consulting other large repositories of temporal Web data, such as Common Crawl, or by simply contacting the Internet Archive and requesting for all domains in the .yu domain. However, unlike comparable archival studies of national domains that had full access to datasets from the Internet Archive or national libraries (Brügger 2015; Hale et al., 2014; Huurdeman et al., 2015), this study's primary goal is to reconstruct a portion of the Web that no longer exists and whose archived locations are unknown. The use of lists as starting points was deliberately chosen to assume an end-user approach to studying the history of the domain, thereby examining the utility of the archived Web—accessed through the IAWM—as a primary source that suffices to narrate the history of the Web, without the

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

aid of secondary, offline sources. The reliance on lists as authoritative sources also demonstrates the contingency and multiplicity of the Web's pasts. If I had only had access to *PC Press* magazine as a source list, the reconstruction of the .yu domain would have been limited to the early years of the domain, depicting primarily the magazine's editorial preference for websites about tourism, technology, e-commerce and news. The capture of Google's search results on the last day the .yu domain existed on the live Web would have generated a history of the last phase of the domain, primarily comprising commercial domains in the midst of changing their suffixes to the domains of Serbia and Montenegro. Between the two extremes, Wikipedia's robustness and diversity as a source list contributed to the generation of the core of the archived Web space. This is surprising given the historical debates around the reliability of Wikipedia as source of knowledge (Giles, 2005). At the same time it is unsurprising, given that Wikipedia is both an archive in itself and a referencing device. Thus, the benefit of using lists as a primary source lies in their distinctiveness, which can further characterize or contextualize archival content.

The missing contextualization of archived data is apparent in the analysis of the changes to the dataset's networked structure over time, which could not have been interpreted without the aid of secondary sources. These indicated that the development of the domain stagnated due both to economic sanctions and to the limits on domain registry while the domain was managed on a voluntary basis and without governmental

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

support. Therefore, historical network analysis alone does not suffice to provide the rich contextual information necessary for understanding the history of the national Web in its full complexity. While the missing context may be found in the content of the archived websites, the spatial reconstruction of the evolution of the networked space adds an important element to the practice of Web historiography as it addresses the missing networked context of the archives' single-document approach.

The tension between large-scale analyses of the spatial contours of the domain and qualitative analyses of smaller sets of websites is apparent when comparing the finding of this study with previous studies on the Yugoslav Internet. Mazzucchelli (2012) found that despite the disintegration of the country, the Yugoslav socio-cultural space is found on the (live) Web in diasporic networks, and Pogačar (2011) showed that the memory of Yugoslavia is preserved on music blogs. Interestingly, the majority of the websites mentioned in these studies are hosted in generic domains such as .net and .org., which, counter-intuitively perhaps, confirm the hypothesis and the findings of this study. While the composition of hosts curated by PC Press magazine and the reconstructed hyperlink networks of the .yu domain between 1996 and 2000 confirm that the Web presence of former Yugoslavia under Milosevic was primarily located in generic Websites, the domain's networked structure after 2000, its archival coverage and migration patterns show that the formal structure of the Web around ccTLDs of sovereign states did not allow for the .yu space to remain the domain of past

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

Yugoslavia, but gradually molded it as the de facto national Web of Serbia. Further studies could examine the outlinking patterns from the other former Yugoslav republics to the .yu domain, as well as compare the reconstruction of the .yu domain with other retired ccTLDs, in order to assess the potential loss of the Web histories of countries currently under transition, such as Syria.

Conclusions

‘Data trash’, argues Wolfgang Ernst, is ‘positively, the future ground for media-anarchaeological excavations’ (2012:120). In many ways, the removal of the .yu domain from the DNS has turned Yugoslavia’s past presence on the Web into ‘data trash’, which this study attempted to excavate. The excavation of that which has been deleted shows the importance of Web archives as the Web’s memory organs and, at the same time, their limits as primary sources for historiographical research. The dependence of the archive on the live Web, in particular, indicates that the Web’s history is best remembered when it is fresh. The longer the distance between the live website and its archived snapshot, the greater the need for the Web historian to use secondary sources for contextualization.

The choice of a deleted national domain as the case study of this research also sheds light on the impact of domain politics on the shaping of national histories on the Web. This study has shown that the possibility of utilizing (archived) Web resources for historiographical research is dictated by the politics of the DNS, which binds together

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

the notions of permanence and the nation-state. The analyses presented in this paper demonstrate the gradual transformation of the .yu domain from a stagnated space of a disintegrated, war-torn area into an emerging national Web space of new sovereign states. The Internet's structural preference of the nation-state thus facilitates the history of sovereign states, but overwrites their unstable pasts.

Acknowledgements

I would like to express my sincere gratitude to Ivan Manojlović from the Museum of Yugoslav History for providing background documentation, to Nikola Smolenski for contributing the seed lists, and to Adam Amram for his valuable research assistance.

¹ The IANA report on the removal of the domain notes that in June 2009 Google indexed 6.2 million pages within .YU (IANA 2010). The Google index of the results that appears in Smolenski's list from March 2000 is 2.5 million pages; However, since Google does not provide access beyond the 1000th result, Smolenski's list comprises the top 1000 results per sub domain (.yu, co.yu, org.yu, ac.yu, .gov.yu).

² To date, most national Web archives provide access to their pages using the open source version of the Wayback Machine. See <http://www.netpreserve.org/openwayback>.

³ Kosovo is an exception. Since to date Kosovo is not officially recognized by the United Nations, it was never delegated a ccTLD.

⁴ This heuristic normalizes the selection of the version of the archived snapshot for the historical network analysis, ensuring that links from the same years have been archived around the same time. (Weltevrede and Helmond, 2012; Huurdeman et al. 2015).

⁵ According to IANA the migration rates are slightly higher. It is reported that as of June 2009 only 200 out of the 4266 remaining .yu domains did not have a match in the .rs domain (IANA, 2010).

Please note: This is the accepted version of the article published online before print, April 28 2016, *New Media & Society*, doi: 10.1177/1461444816643790

References

- Baeza-Yates R, Castillo C and Efthimiadis EN (2007) Characterization of national Web domains. *ACM Transactions on Internet Technology (TOIT)* 7(2): 9.
- Ben-David A and Huurdeman HC (2014) Web archive search as research: Methodological and theoretical implications. *Alexandria* 25(1): 93–111.
- Bilbija B (2013) Dijaspóra može da promeni Srbiju (Diaspora can change Serbia). *Politika*, December 29 (accessed 31 May 2015) <https://archive.is/T8RLq>
- Brügger N (2009) Website history and the website as an object of study. *New Media & Society* 11(1-2): 115–32.
- Brügger N (ed) (2010) *Web History*. New York: Peter Lang Publishing.
- Brügger N (2012) Historical network analysis of the Web. *Social Science Computer Review* 00: 1–00 (accessed 31 May 2015) doi: [10.1177/0894439312454267](https://doi.org/10.1177/0894439312454267).
- Brügger N (2013) Web historiography and Internet studies: Challenges and perspectives. *New Media & Society* 15(5): 752–64.
- Brügger, N. (2015, in press). Probing a nation's web sphere: A new approach to web history and a new kind of historical source. In Gerard Goggin & Mark McLelland (Eds.), *The Routledge Companion to Global Internet Histories*.
- Chun WHK (2008) The enduring ephemeral, or the future is a memory. *Critical Inquiry* 35(1): 148–71.
- Chun WHK (2011) *Programmed Visions: Software and Memory*. Cambridge: MIT Press.
- Diminescu D (2012) Digital methods for the exploration, analysis and mapping of E-Diasporas. *Social Science Information* 51(4): 451–458.
- Ernst W (2012) *Digital Memory and the Archive*. Minneapolis: University of Minnesota Press.

Please note: This is the accepted version of the article published online before print, April 28 2016, New Media & Society, doi: 10.1177/1461444816643790

Gakovic V and Szymczyk K (2007) Serbian Domain Name System Reborn. *Managing Intellectual Property: The Global IP Resource* 166: 36–40 (accessed 31 May 2015) http://heinonline.org/HOL/Page?handle=hein.journals/manintpr166&g_sent=1&id=38.

Giles J (2005) Internet encyclopaedias go head to head. *Nature* 438(7070): 900–1.

Gomes D and Silva MJ (2005) Characterizing a national community web. *ACM Transactions on Internet Technology (TOIT)* 5(3): 508–31.

Halavais A (2000) National borders on the World Wide Web. *New Media & Society* 2(1): 7–28.

Hale SA, Yasseri T, Cowls J, Meyer ET, et al. (2014, June) Mapping the UK webspace: Fifteen years of British universities on the web. In *Proceedings of the 2014 ACM Conference on Web Science*. ACM, 62–70 (accessed 31 May 2015) [arXiv:1405.2856](https://arxiv.org/abs/1405.2856).

Hockx-Yu H (2014) Access and scholarly use of web archives. *Alexandria* 25(1): 113–27.

Huurdeeman HC, Kamps J, Samar T, de Vries AP et al. (2015) Lost but not forgotten: Finding pages on the unarchived web. *International Journal on Digital Libraries* 16(3): 247–265.

Internet Assigning Numbers Authority (IANA) (2001) *Report on deletion of the .zr Top-Level Domain* (accessed 31 May 2015) <http://www.iana.org/reports/2001/zr-report-20jun01.html>.

Internet Assigning Numbers Authority (IANA) (2010) *Removal of the .YU domain formerly representing Yugoslavia* (accessed 31 May 2015) <https://web.archive.org/web/20101205175937/http://www.iana.org/reports/2010/you-report-01apr2010.html>.

Internet Corporation for Assigned Names and Numbers (ICANN) (2006) Discussion paper on retiring Country Code Top Level Domains (accessed 31 May 2015) <https://www.icann.org/news/announcement-2-2006-12-05-en>.

Please note: This is the accepted version of the article published online before print, April 28 2016, New Media & Society, doi: 10.1177/1461444816643790

Keenan T (2001) Looking like flames and falling like stars: Kosovo, 'the First Internet War'. *Social Identities* 7(4): 539–550.

Maguire M (2003) East Timor's .TP: From a virtual initiative to a political reality. In Wass ES (ed) *Addressing the World: National Identity and Internet Country Code Domains*. Lanham, MD: Rowman & Littlefield Publishers: 17–30.

Manojlović I (2014) The Museum of Yugoslav History: The acquisition of the .yu domain. In Domanovic A, *From Yu to Me*. Firstsite.

Masanès J (2005) Web archiving methods and approaches: A comparative study. *Library Trends* 54(1): 72–90.

Mazzucchelli F (2012) What Remains of Yugoslavia? From the geopolitical space of Yugoslavia to the virtual space of the Web Yugoslosphere. *Social Science Information* 51(4): 631–48.

Mueller M (2004) *Ruling the Root: Internet Governance and the Taming of Cyberspace*. Cambridge: MIT Press.

Phillips ME (2005) What should we preserve? The question for heritage libraries in a digital world. *Library Trends* 54(1): 57–71.

Pogačar M (2010) Memonautica: Online Representations of the Yugoslav past. In Riha D (ed) *Humanity in Cybernetic Environments*. Oxford, UK: Inter-Disciplinary Press, 23–34.

Rogers R (2002) Operating issue networks on the web. *Science as Culture* 11(2): 191–213.

Rogers R (2013) *Digital Methods*. Cambridge: MIT press.

Schneider SM and Foot KA (2005) Web sphere analysis: An approach to studying online action. In Hine C (ed) *Virtual Methods: Issues in Social Research on the Internet*. Oxford, UK: Berg, 157–170.

Please note: This is the accepted version of the article published online before print, April 28 2016, New Media & Society, doi: 10.1177/1461444816643790

Smolenski N (2009) YU domain transition. *Wikibots-I Digest* 27(1) 17 February (accessed 31 May 2015) <https://archive.today/1YXAv>.

Thelwall M and Vaughan L (2004) A fair history of the Web? Examining country balance in the Internet Archive. *Library & Information Science Research* 26(2): 162–76.

Wass ES (ed) (2003) *Addressing the World: National Identity and Internet Country Code Domains*. Lanham, MD: Rowman & Littlefield Publishers.

Weltevrede E and Helmond A (2012) Where do bloggers blog? Platform transitions within the historical Dutch blogosphere. *First Monday* 17(2–6): 00–00 (accessed 31 May 2015) <http://firstmonday.org/ojs/index.php/fm/article/viewArticle/3775>.

Anat Ben-David is a lecturer in the department of Sociology, Political Science and Communication, the Open University of Israel. Her research interests include Internet geopolitics, Web historiography and digital methods for Web research.

Tables

List Name	Time-Range	Source Description	No. of .yu URLs	No. of Hosts	No. of Hosts unique to this list
Google	2010	A snapshot of Google search results for the .yu domain and its sub-domains one day before the domain was removed from the Internet (Source: Nikola Smolenski).	4761	1324	947
EUNET	2009	A list of .yu URLs hosted by the Serbian ISP EUNET in 2009 (Source: Nikola Smolenski).	1361	393	284
Wikipedia	2010	List of all .yu URLs linked to by all Wikipedia pages in 2009 before the domain removal (Source and Script: Nikola Smolenski).	3441	2827	2419
PC Press	1997-2008	A yearly list of the top 50 most popular .yu websites curated and ranked by computer magazine <i>PC Press</i> between 1997-2008 (Source: Museum of Yugoslav History / Internet Archive).	400	266	190

Table 1. Description of the Source Lists

Total			10333	4810	3840
--------------	--	--	--------------	-------------	-------------

List Name Sub-domain	EUNET	Google	PC Press	Wikipedia
Ac.yu	0	49	0	155
Cg.yu	0	0	1	185
Co.yu	263	603	55	1036
Com	0	4	66	0
Edu.yu	2	56	0	120
Gov.yu	0	28	0	88
Net	0	1	12	0
Org	0	4	11	0
Org.yu	19	165	3	788
Other	0	14	15	47
Rs	0	23	27	0
Total	284	947	190	2419

Table 2. The sub-domain distribution of unique hosts, per source list

TLD	No. of Unique Hosts	Examples
Com	65	nostalgija.com; beograd.com
Org	13	serbiatraveles.org; freeserbia.org;
Net	12	beotel.net; b92.net;
Info	3	putovania.info; stetoskop.info

Table 3. Unique Yugoslav hosts outside the .yu domain, extracted from the *PC Press* list

Figures

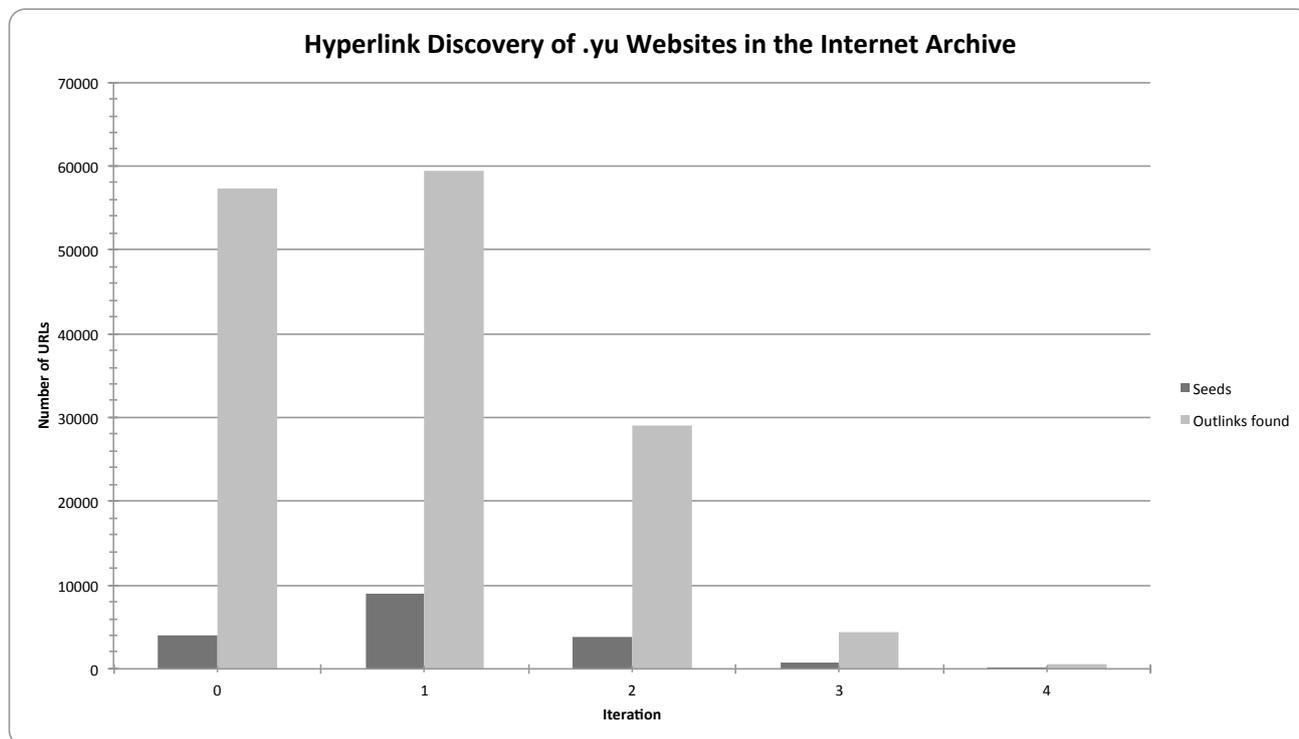


Figure 1. The number of unique .yu URLs found using the hyperlink discovery method, per iteration

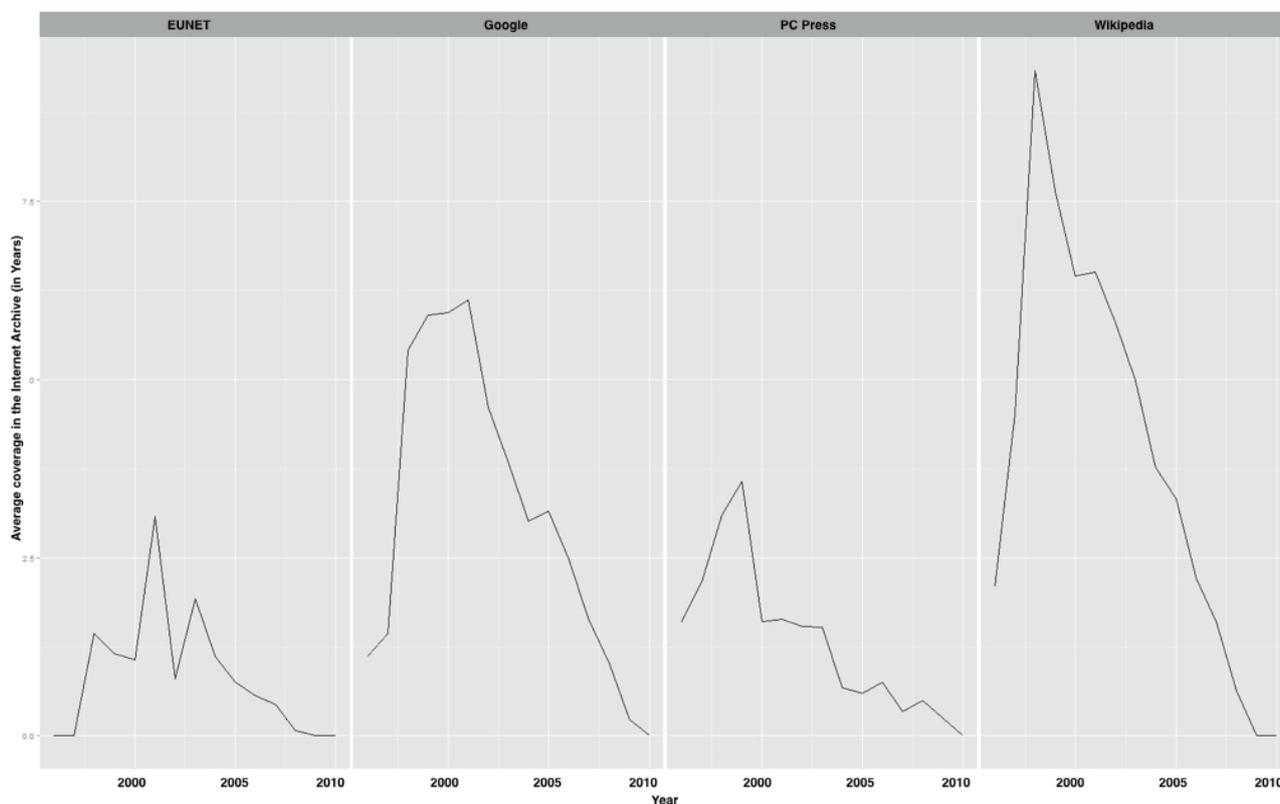


Figure 2. The average number of years of the Internet Archive's coverage of unique sources from each list (1997-2010). The number of years is calculated as the distance between the timestamp of the first and last archived snapshot of each unique host

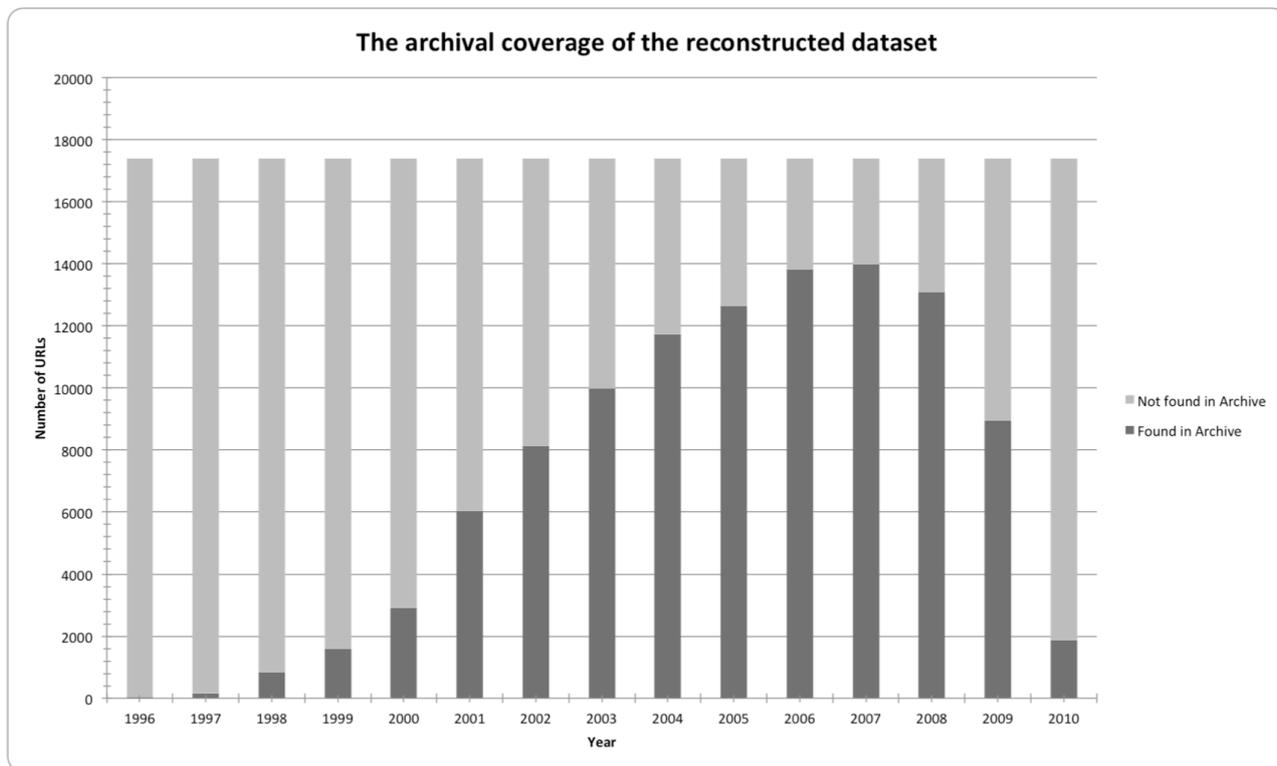


Figure 3. The accessibility of .yu websites in the Internet Archive over time

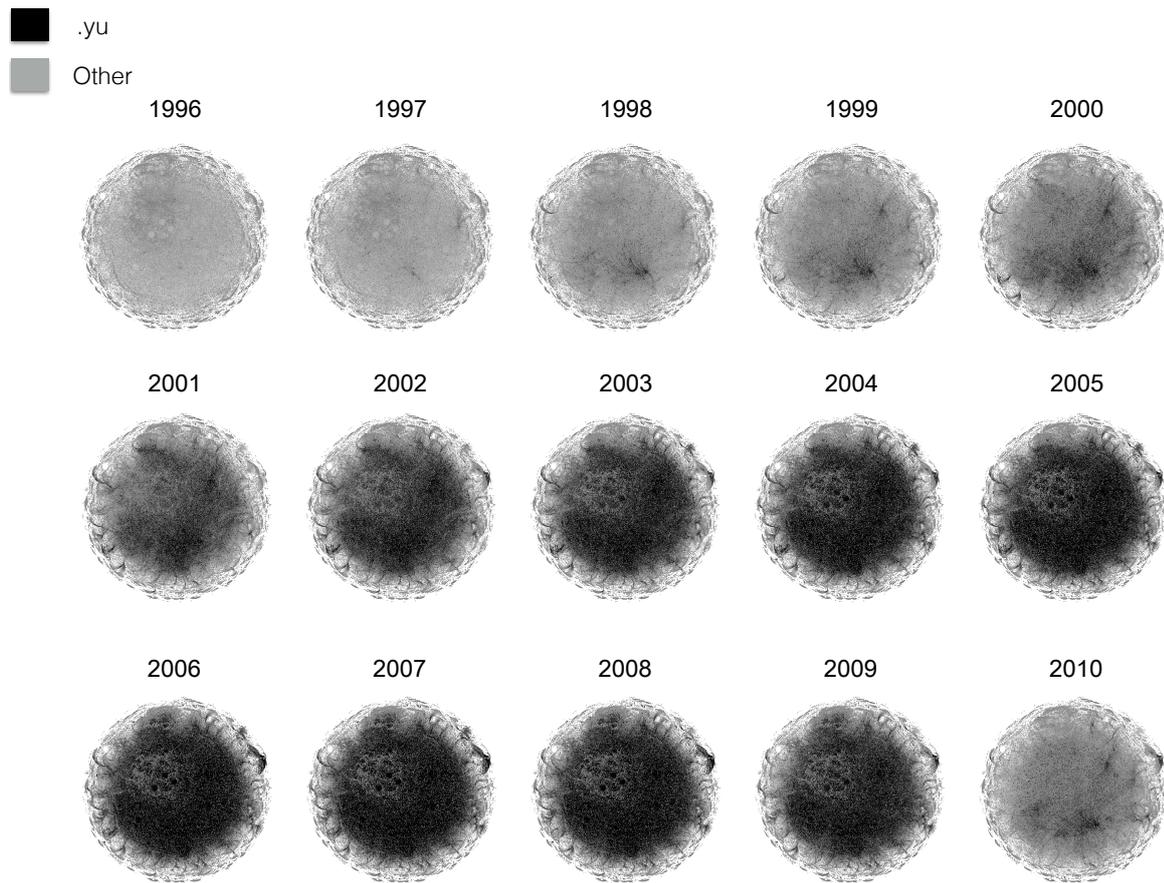


Figure 4. The evolution of the .yu networked space over time.

The distribution of outlinks from the .yu domain, 1997-2010

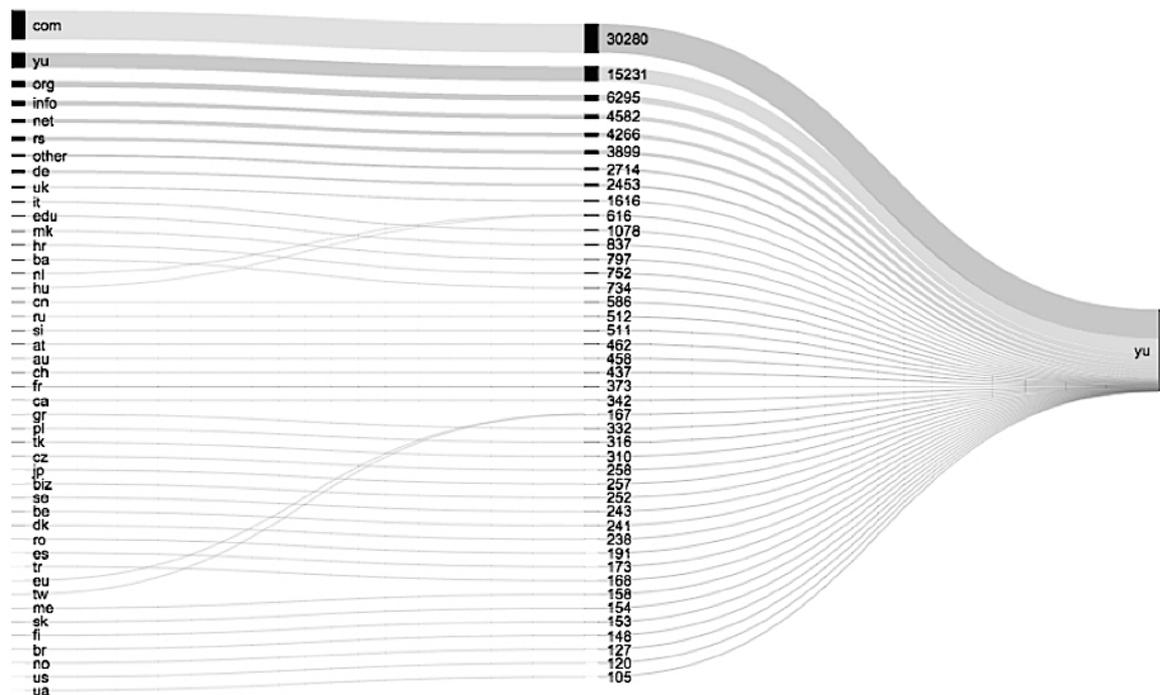


Figure 5. The distribution of outlinks from the .yu domain to other TLDs