



## Article

# Investigating the Country of Origin and the Role of the .eu TLD in External Trade of European Union Member States

Andreas Giannakoulopoulos <sup>1,\*</sup> , Minas Pergantis <sup>1,\*</sup> , Laida Limniati <sup>2</sup> and Alexandros Kouretsis <sup>1</sup>

<sup>1</sup> Department of Audio and Visual Arts, Ionian University, 7 Tsirigoti Square, 49100 Kerkira, Greece; akourets@gmail.com

<sup>2</sup> BrilliantPR Digital Agency, 340 Kifisias Str., 15451 Athens, Greece; laida.limniati@brilliantpr.gr

\* Correspondence: agiannak@ionio.gr (A.G.); a19perg6@ionio.gr (M.P.)

**Abstract:** The Internet, and specifically the World Wide Web, has always been a useful tool in the effort to achieve more outward-looking economies. The launch of the .eu TLD (top-level domain) in December of 2005 introduced the concept of a pan-European Internet identity that aimed to enhance the status of European citizens and businesses on the global Web. In this study, the countries of origin of websites that choose to use the .eu TLD are investigated and the reasoning behind that choice, as well as its relation to each country's economy and external trade are discussed. Using the Web as a tool, information regarding a vast number of existing .eu websites was collected, through means of Web data extraction, and this information was analyzed and processed by a detailed algorithm that produced results concerning each website's most probable country of origin based on a multitude of factors. This acquired knowledge was then used to investigate relations with each member-state's presence in its local ccTLD, its GDP and its external trade revenue. The study establishes a correlation between presence in the .eu TLD and external trade that is both independent of a country's GDP and stronger than the relation between its local ccTLD presence and external trade.

**Keywords:** internet domains; web data extraction; e-commerce; international trade; European identity; big data; TLDs; World Wide Web



**Citation:** Giannakoulopoulos, A.; Pergantis, M.; Limniati, L.; Kouretsis, A. Investigating the Country of Origin and the Role of the .eu TLD in External Trade of European Union Member States. *Future Internet* **2022**, *14*, 174. <https://doi.org/10.3390/fi14060174>

Academic Editor: Michael Sheng

Received: 19 May 2022

Accepted: 2 June 2022

Published: 4 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Internet is frequently viewed as a tool for globalization and the elimination of borders. In fact, Ershov argues that if there is a free flow of information across domestic boundaries and anyone, regardless of location, can join at any time, then nationalism should decline, as its foundation is shaken by the discourse of new media [1]. In reality this might not always be the case, as the Internet increasingly reproduces features of traditional media and state authorities exert control over it [1]. Digital information and mediated communication predominate in today's culture.

The modern online media landscape is comprised of web versions of traditional media as well as new online native media and social networks. Content generation and transmission are no longer limited to large organizations; anyone can frequently upload information in multiple formats that can be easily updated [2] independently of the location. In this way, the Internet, is a quite democratic medium, since it offers a platform for all voices to be heard, and at the same time unites people from different countries and cultures.

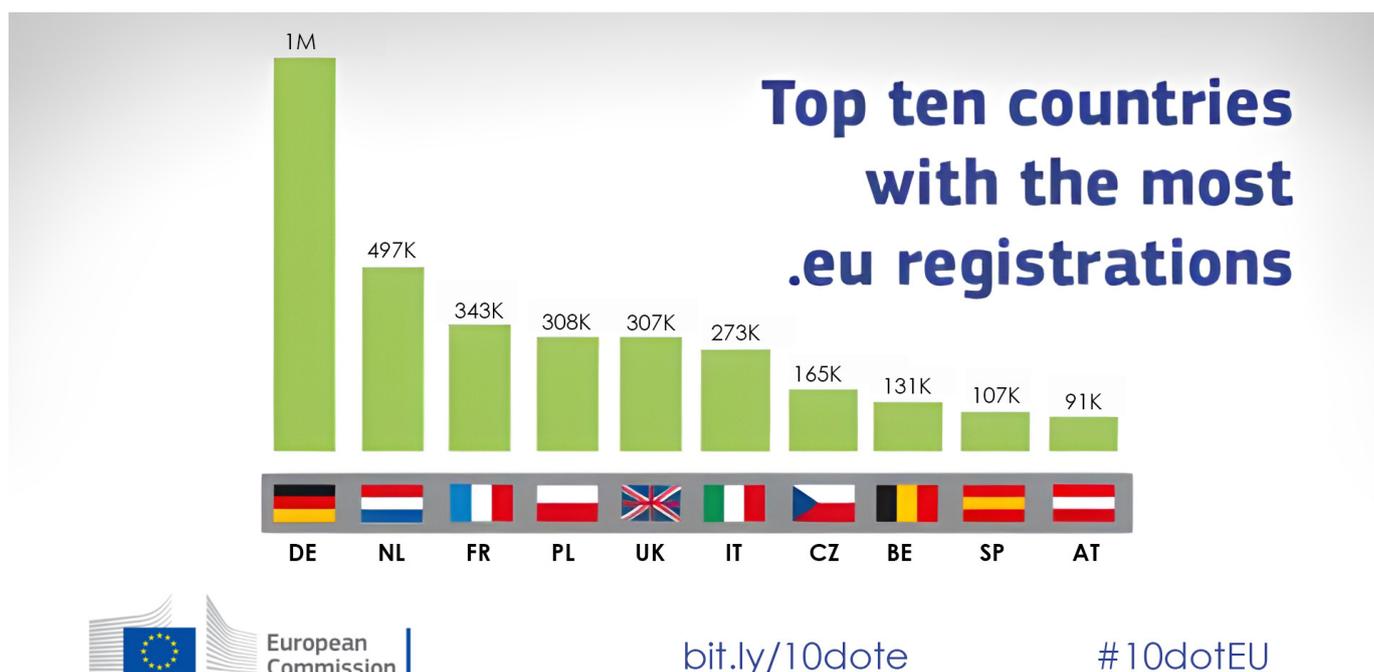
This study tries to investigate and evaluate the usage of the .eu top-level domain (TLD). The .eu is the European Union's TLD and in contrast to other existing country code top-level domains (ccTLD), it represents the whole European Union and not a country itself. However, it falls under the same category with other ccTLDs such as .gr, .es, .de, etc., contrasting with the .africa domain that represents the African Union and is a TLD that it is generic according the Internet Assigned Numbers Authority (IANA) [3]. The .eu top-level domain can be perceived as a way for organizations operating in the European Union to

reflect their European identity in cyberspace [4]. From its conception it was destined to generate interest as the Internet label for Europe [5]. As stated clearly in the regulation on the .eu implementation, the .eu TLD was created with the objective of accelerating electronic commerce in the e-Europe initiative and to promote the use of, and access to, Internet networks and the virtual marketplace based on the Internet [6]. Through this, it enables users to establish a pan-European Internet identity for their websites and emails. It is accessible to all EU-based businesses and organizations, as well as all EU citizens and permanent residents [7].

In general, the domain name system has become more than just a technological convention that appears as a suffix at the end of an Internet address. It has the ability to affect social change and incorporate national identities and priorities thanks to the use of ccTLDs. It is a means of communicating cultural values [8]. Postel and his team, who founded the Internet Assigned Numbers Authority at the end of the 1980s, anticipated an endless debate about what a nation consisted of and therefore what should or should not be included in the list of codes since they did not want to end up as referees in political debate [8]. Although Postel thought of domains as country zip codes, they began to have more political and social meaning [8]. Gallup & Gandhi [8], almost 20 years ago, tried to shed light on the choice of domains of two different countries, India and Chile. Chile with a strong national identity chose the ccTLD .cl while India, at that time with a less strong national identity, was not fond of the .in ccTLD, instead they were in favor of more generic domains such as .com [9]. Patrik Linden mentions a correlation between the Islamic world and the generic global TLDs, since it represents a more progressive approach than choosing their national domains [8]. Another interesting example by Wass [8] comes from the USA, where .com might be considered as USA's ccTLD, but in fact it stands for commercial, and the official ccTLD is .us.

Ørmen's et al. recent study on the Internet use in Europe, China, and the USA indicates a convergence between regions, with sociodemographic and also national and regional differences. Furthermore, in terms of mass and networked communication, the youngest age groups share more similarities with each other across countries than with the oldest age groups within the same country [9].

Registering the .eu domain corresponds to establishing an Internet-based digital European identity and also a pan-European identity or presence, which definitely helps in the case of e-commerce. The .eu TLD should help the European Union's internal market achieve greater visibility in the online marketplace of the Internet. It is important that the .eu TLD domain provides a link that can be easily established with the European Community, its associated legal framework, and the European marketplace [6]. Individuals or organizations with the .eu domain project the message that they are from the European Union, emphasizing their credibility, reliability, and trustworthiness, thereby boosting user confidence [10]. Furthermore, the .eu domain could be chosen for technical reasons, such as availability (if a local domain or .com has already been taken), it offers flexibility and is suitable for local, national, or international regions [10]. Another reason might also be financial, since the .eu domain seems to be the most economical as it is suggested by some registration companies and it is also suggested as a domain to choose if you want to show that your business is available through Europe. In this case, it will not come as a surprise to find that the majority of the websites using the .eu domain are either European organizations or e-shops, enterprises etc. In fact, Wahdani and Alfaouri [4], mention that the EU has established its own top-level domain (.eu), paving the way for later implementation of the free movement of services via the Single Digital Market (SDM). It is true that an increasing number of Europe-focused businesses enhance their online presence by using the .eu domain extension [11]. According to the EC "At the end of 2015 .eu was the 11th largest TLD in the world" [11] and its usage was embraced by businesses of the EU member states, with Germany, the Netherlands, France, and Poland leading in the registrations as seen in Figure 1.



**Figure 1.** Top ten countries with the most .eu registrations in 2015 [11].

The .eu top level domain went officially live on 7 December 2005, although it had been technically functional for some time prior [12]. The desire to have an online identifier representative of the EU itself, similar to how each country has its own ccTLD, was a primary motivation. The anticipated prevalence of European domain names was expected to contribute to raising awareness of the EU and possibly popularizing the concept of the bloc as a truly unified territory [12]. At the same time, it is becoming more and more evident that the .eu TLD creation had, and continues to have, an e-commerce orientation, since recently the EURid and the E-commerce Foundation announced the strengthening of their collaboration [13].

In general, the European Union, in the process of becoming an ‘imagined community’, tries to awaken a European consciousness with initiatives by promoting its symbols, while respecting the content of national cultures [14]. The .eu TLD can be also perceived as such an initiative in the digital space. Sassatelli [14], mentions that the idea of ‘Europe’ as the basis of an identity is fostered by the EU’s search for means of legitimization.

Considering that national identity discussions are increasing, we must have a solid understanding of what national identity is and how it relates to EU support [15]. The notion of a European identity was discussed way before the aforementioned initiatives and studies. Anthony Smith [16] investigates the relationship between political consciousness and European political identity and wonders whether a European identity is possible. Later, Carey [17], attempts to explain the behavior of EU citizens based on their national identities. It defines national identity in three distinct ways but argues that European Identity will always be inferior to national identity. Benedict Anderson [18] assumes that nations are mental constructions, or “imagined communities”, that nationalized political subjects perceive as distinct political entities in order to promote an agenda. Since this study is not focused on political aspects of the national identity, there is no need to enter into further details about Anderson’s view on that, but it can be safely assumed that all nations are in some form imagined communities. Adopting the Wodak [19] team’s assumptions, it becomes clear that ‘national identity’ refers to a mix of similar perceptions and conceptual frameworks, emotional behaviors and attitudes, and behavioral conventions that bearers of this ‘national identity’ share collectively and have internalized through the socialization

process [19]. As distinct forms of social identities “are produced, reproduced, transformed, and dismantled” [19].

Concerning European identity, Cram [20] in her effort to comprehend the role identity plays in the European integration process, distinguishes three basic categories: a self-allocated label or role (identify as European), a state of being (“I agree to some extent with EU and/or its results”) and a political behaviour (“I support EU”). Bruter [21] makes a distinction between the cultural and civic aspects of European identity. According to his study, “many respondents identify with Europe and the EU, mostly in civic terms”.

Studies are divided on whether national identity strengthens or weakens EU support [15]. According to Fligstein [22] “Europeanization can be defined as the development of new social arenas in which groups (such as states, nonprofit organizations, individuals, or businesses) from more than two countries regularly interact”. If we take into account that Europeanization is a limited aspect of globalization, and try to investigate the relationship between globalization and national identity we can also see that the studies are also puzzling. On the one hand, there are studies suggesting that globalization reduces national identity, other studies suggest the opposite [23]. According to Pew Research Center [24], the views of European Union over time have leaned on to the positive direction. For instance, Greece has seen a 26% surge in favorable views of the EU from 2016 to 2019. This does not necessarily mean of course, that they identify more as Europeans than Greeks for example, rather that they think it is beneficial to be a part of this Union. Europeanization adds to, and supports, the decision of European businesses to conduct business via the .eu TLD, utilizing European infrastructure and Internet venues [25].

Another study from Eurobarometer on Values and Identities of European citizens suggests that, at a personal level, the important identities of EU citizens are their family (81%) and national identity (73%). Despite their strong regional and national identity, the same study showed that 56% of respondents across the EU indicate identifying with being European, 28% are noncommittal, and only a minority of 14% indicate not identifying with being European.

More specifically, citizens from Hungary (76%), Slovakia (75%), Malta (72%), Cyprus and Poland (both 67%), Romania and Czechia (both 66%), Spain and Slovenia (65%), Italy (64%), and Lithuania, Latvia, and Austria (all 63%) identify with being European. On the contrary, respondents from Greece and Sweden (both 42%), Croatia (45%), Belgium and Estonia (both 46%), the Netherlands (48%), and Finland (49%) are the least likely to identify with being European. At the same time, younger people (15–24 years old) are slightly less likely to identify with being European than those aged 55 years old or older. (54% in comparison to 59%) [26]. This comes in a way in agreement to Ørmen’s team study where younger people share more similarities with each other across countries [9]. Furthermore, according to the study, people in rural villages are slightly less likely (55%) to identify with being European than those living in big cities (60%). Figure 2 presents the detailed graph concerning how different EU citizens identify as European.

From the above literature review, we can assume that in general, the sentiment towards Europe is steadily on the rise, but that does not mean that it surpasses a national identity. Identifying as European does not constitute denial of a citizen’s national identity, but rather indicates that someone wants to gain the benefits of being European. Hence, this might be closer to Bruter’s distinction between cultural and civic aspects of European identity. The same can be applied to the ccTLDs sphere. By choosing the .eu domain, it does not mean that someone identifies as more European, but that they want to take advantage of the civic aspect the European identity offers, whether these are profit-related, location related, etc.

E-commerce can be perceived as such a profit-related sector. In fact, one-third of .eu TLD domain name owners are involved in business [25], and it would be informative to investigate how they perceive the EU and the EU’s economic sustainability, as well as how necessary (and worthwhile) it is for them to advertise their European identity. According to reports, the .eu TLD is an instrument of European identity that does not damage national

registrations, i.e., a rise in domain name registrations inside the .eu TLD does not result in a decline in registrations within the ccTLD of the member states (.de, .nl, etc.) [25].

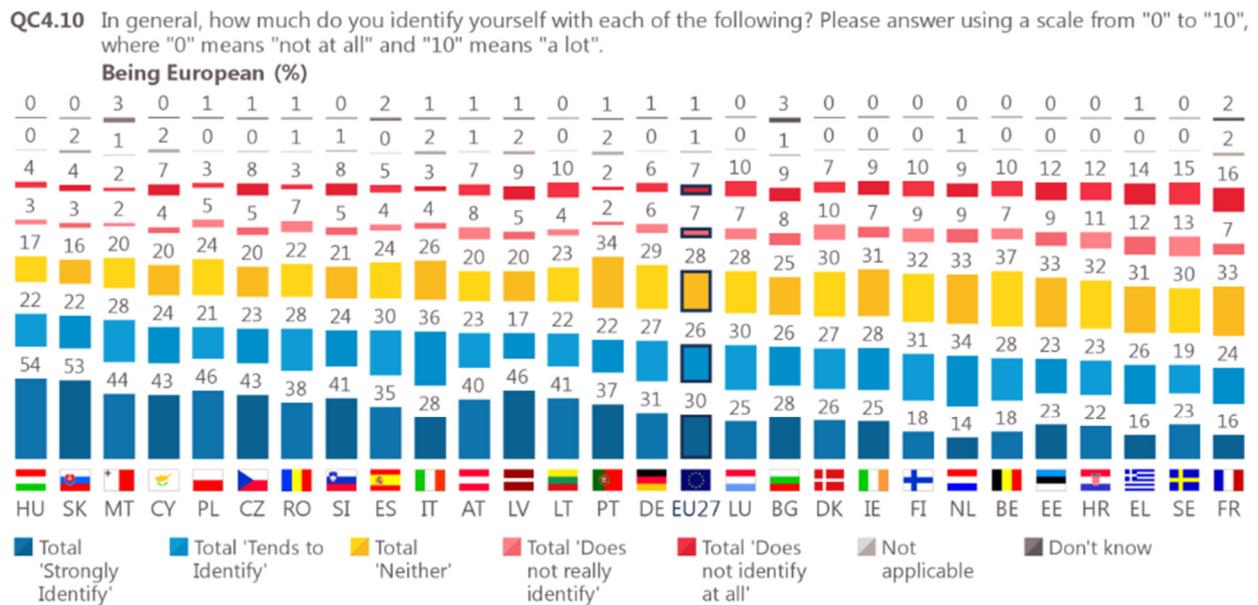


Figure 2. Identifying as European according to the special Eurobarometer 508 [26].

Inspired by the above, this study attempts to practically map the actual landscape of each country's .eu TLD presence through investigating contemporary websites that are not just registered but actually online and operational. It then proceeds to compare that presence with each country's economic size as measured by its GDP and contrast it with the country's Web presence in the national ccTLD landscape of the European Internet. Furthermore, the study investigates the interrelation between these presences and each country's external trade as indicated through exports to countries outside the European Union.

The study presents in great detail a method of collecting data from a large number of websites and inferring information regarding their country of origin. This methodology can be expanded and with alterations used to serve other purposes beyond the scope of this research. Moreover, the landscape of the .eu TLD usage by EU member state as it is surmised through investigating actual existing public websites provides a valuable glimpse regarding the adoption of the common European TLD by each country. In addition to that, this usage (expressed as a percentage of each member state's GDP) was found to be closely related but not identical to the usage of each country's ccTLD, implying different factors that might influence it. Finally, by investigating the relationship between .eu usage by GDP and the ratio of external trade outside the EU (Extra-EU) by GDP, a significant positive Spearman correlation is established, which indicates a monotonic relationship between these two input variables. In simple terms, an increase in a country's Extra-EU trade by GDP ratio indicates an increase in a country's presence in the .eu TLD landscape, thus connecting .eu presence with the idea of market openness.

For the purposes of this research, the United Kingdom has been manipulated in all algorithms, metrics, and theoretical approaches as no longer being a member of the Union because of its official secession on 1 February 2020. It should be noted that the UK's official involvement in the .eu TLD has officially ended as of early 2022 and all websites that could not be linked to a different member state were withdrawn by the EURid agency [27].

## 2. Methodology

In order to map the presence of each European Union member state in the landscape of .eu TLD usage the study proceeds to:

- Collect a multitude of websites in the .eu TLD
- Collect information from a representative sample of these websites using Web data extraction techniques.
- Algorithmically evaluate that information to discover each website's country of origin. Each step of this process is thoroughly detailed in this methodology section.

### 2.1. Website Collection

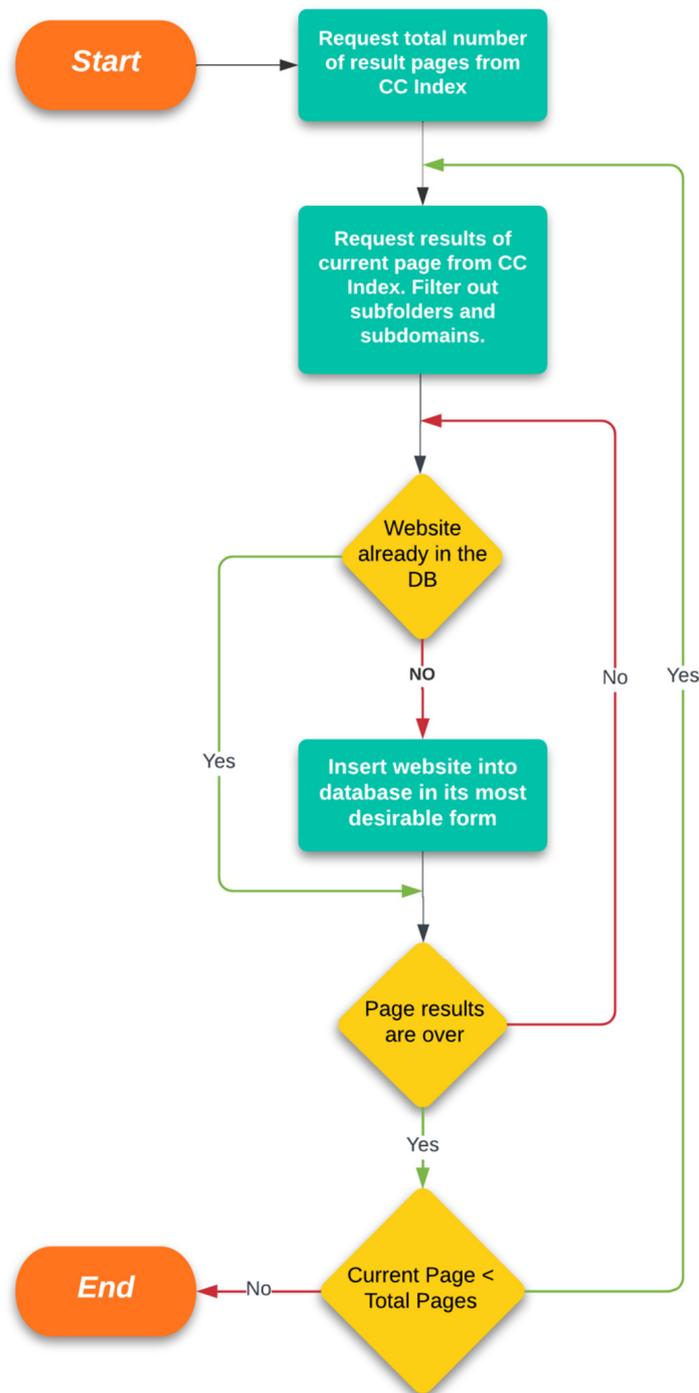
In order to collect and process data, multiple algorithms were used that will be further detailed in this section. All algorithms were developed using the PHP general purpose scripting language and were run on an Ubuntu distribution of the Linux operating system. The cURL PHP library was used to preform HTTP requests when necessary. All information collected was stored in a relational database using the MariaDB database management system. This specific technology stack was purposefully selected due the popularity of its components, which ensured extended functionality support in the form of libraries, a high frequency of updates and of course compatibility and strict adherence to Web standards. The researchers' familiarity with these technologies also played an important role in this selection.

The first step in studying websites that are using the .eu Top Level Domain (TLD) was acquiring an accurate and up-to-date list of such websites. To achieve this, an algorithm was developed that queried the Common Crawl Index servers for Web pages and collected all the unique domains found. Common Crawl is a non-profit foundation providing access to web information through an open database of web crawl data [28]. The data provided by Common Crawl was indexed in January 2022 and was the latest available at the time the study was conducted. The algorithm queried the Common Crawl Index servers for the number of pages of results using the .eu TLD and then proceeded to request each page, filter out sub-folders and subdomains while retaining the unique second level domains (SLD), detect their protocol (http or https) and usage of the www subdomain and record their full URL and the website language as provided by CC Index in a database table. This process yielded a total of 203,020 unique domains belonging to the .eu TLD. Figure 3 depicts a flowchart of this process.

### 2.2. Web Data Extraction

With the domains collected, the task at hand was to infer the country more closely related to each website. In order to reach a trustworthy conclusion concerning the main country of origin of a website, a data series was collected regarding the various aspects of the website's online presence. The main tool for collecting this information was Web data extraction (or Web scraping). Table 1 presents the various variables that the Web data extraction algorithm recorded, with their types, multiplicity, and a description of each variable. These variables were selected based on the feasibility of their collection and their relevance to the goal of establishing a website's country of origin and are based around both technical characteristics of the hosting process and website content in terms of hyperlinks and language usage.

In order to detect the country of the host of the website (*server\_country*) the algorithm makes use of the GeoLite2 PHP library by MaxMind with the latest database update. The GeoLite2 databases are free IP geolocation databases provided by MaxMIND [29] and were used through an implementation of their PHP API [30]. Other country code TLDs (ccTLDs) having a registered and functioning domain with the same SLDs were detected through HTTP requests and recorded in the *other\_domain\_countries* variable.



**Figure 3.** Flowchart of the .eu domains discovery algorithm.

To collect the information regarding a website’s registrar (*registrar*, *registrar\_web\_country*, *registrar\_web\_lang*) the algorithm extracted the registrar’s URL from the EURid website. EURid is the regulatory institution responsible for the usage of the .eu TLD as it has been designated by the European Commission. The EURid website offers “whois” functionality that reports the registrar of any .eu website and that registrar’s Web page. After collecting the registrar’s website, the data extraction algorithm recorded its ccTLD if it belonged to an EU member state and proceeded to scrape its home page in order to detect the language used there.

**Table 1.** Variables collected by the data extraction algorithm.

Variable Name	Type	Multiplicity	Short Description
<i>server_country</i>	ISO 3166-1 alpha-2	1	The ISO country code of the IP hosting the website
<i>other_domain_countries</i>	ccTLD	1...n	The ccTLDs of any EU country that has the same SLD as the website.
<i>registrar</i>	URL	1	The URL of the website of the registrar that the website is registered with.
<i>registrar_web_country</i>	ccTLD	1	The ccTLDs of the website of the registrar that the website is registered with
<i>registrar_web_lang</i>	ISO 639-1	1	The language of the website of the registrar that the website is registered with
<i>page_detected_lang</i>	ISO 639-1	1	The language of the home page of the website
<i>mail_countries_str</i>	ccTLD	1...n	The ccTLDs of any EU country that appears in mailto: links in the home page of the website
<i>phone_countries_str</i>	ISO 3166-1 alpha-2	1...n	The ISO country code of any EU country telephone codes that appears in tel: links in the home page of the website
<i>addr_detected_countries_str</i>	ccTLD	1...n	The ccTLDs of any EU country that appears in full text in English in the home page of the website
<i>gmap_countries_str</i>	ccTLD	1...n	The ccTLDs of any EU country that appears in full text in English in a page of the website that contains a map
<i>gmap_languages_str</i>	ISO 639-1	1...n	The language of any page of the website that contains a map
<i>gmap_link</i>	URL	1	The URL of google map that appears on the website
<i>gmap_country</i>	ISO 3166-1 alpha-2	1	The ISO code of any EU country that appears on the above google map link

The language detection was carried out by a subroutine making use of the language-detection PHP library by Patrick Shur [31]. This subroutine analyzed the Web page's DOM structure and collected text strings from the title element, the meta description tag and any text elements in the page's body such as paragraphs, lists, and so on. If the page presented adequate text content (more than 150 characters) the language of that content was inferred and recorded in the database.

Following that, the crawling algorithm proceeded to scrape the homepage of the website being investigated. It used the language detection subroutine as presented above to infer the language used (*page\_detected\_lang*). Additionally, the text elements of the homepage were investigated for any appearance of an EU country name in English and each instance was recorded in the *addr\_detected\_countries\_str* variable.

The next step in data collection was to search the homepage's DOM for all anchor elements and collect their links. If an EU country's ccTLD was detected in a *mailto:* link, that country was added to the *mail\_countries\_str* variable. If an EU country's area telephone code was detected in a *tel:* link, that country was added to the *phone\_countries\_str* variable. The function of detecting the area code and extracting the appropriate country code was performed by a subroutine that made use of the *libphonenumber-for-php* PHP library from Joshua Gigg, that was based on Google's *libphonenumber* Java, C++, and Javascript library [32]. Finally, the algorithm identified any google map links present in the homepage. If such a link was found, it was recorded in the *gmap\_link* variable.

With all homepage links collected, the crawling algorithm proceeded to scrape all internal Web pages that appeared as links in the homepage up to a maximum of 100 pages. The maximum limit was instated to ensure a reasonable maximum amount of time required to investigate a single website. In each of these pages the algorithm also attempted to identify *mailto:* links, *tel:* links and google map links and also extract the relevant information in a similar manner.

For all pages crawled, homepage or otherwise, if a google map link was identified in the DOM, the algorithm detected and recorded that specific page's language and any instance of EU country name in English in the variables *gmap\_languages\_str* and

*gmap\_countries\_str*. The logic behind this was that pages with map links are good candidates for including address information and also present a great chance of containing content in the language that the website’s developers deemed most important. Moreover, if no google map link had been detected in the homepage the present detected link was recorded instead in the *gmap\_link* variable.

Finally, using the *gmap\_link* variable and the API of the Nominatim search engine for OpenStreetMap data [33], a subroutine was created that was able to accurately determine which country the google map link referred to. In order to achieve this, the subroutine extracted the latitude and longitude from the link where possible and made an HTTP request to Nominatim’s API and then retrieved the country code from the API’s response. This information was recorded in the *gmap\_country* variable. Figure 4 presents the algorithm as detailed above in a flowchart.

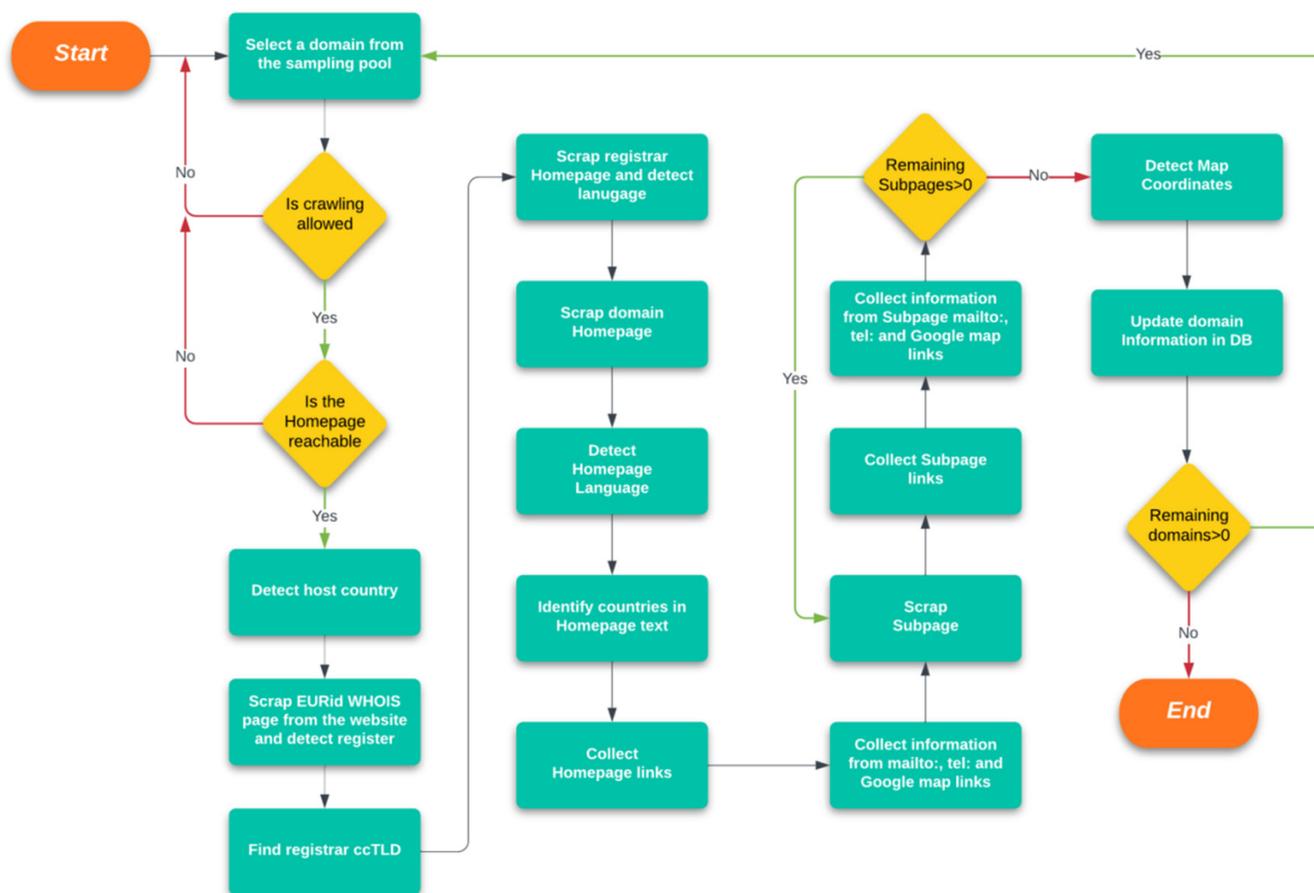


Figure 4. Flowchart of the Web data extraction algorithm.

### 2.3. Country of Origin Evaluation

The cornerstone of this research methodology was the country detection algorithm that was developed. It aimed to process all relevant information collected and recorded in the database by the Web data extraction algorithm with the purpose of inferring the country most closely related to each specific .eu domain.

The country detection algorithm evaluated the information collected on the basis of how indicative each piece of data was to discovering the country of origin of an .eu domain using a point system. Four different weights were devised and assigned a specific point value. These weights are as follows:

- Marginally relevant information: 4 points
- Relevant information: 9 points
- Important information: 19 points

- Very important information: 29 points

The point values of each weight were purposefully selected not to be round numbers in order to minimize the chance for ties. Despite that, ties did still occur and a special tie breaking subroutine that will be detailed further below was developed.

The information regarding where each website is hosted is difficult to evaluate because some EU countries boast a huge number of data-centers while others much less so. That means that when a website is hosted in a country with many data centers it is less relevant than if the website were hosted in a country with fewer. After consulting with the number of data centers per EU country [34], it was decided that the information that a website is hosted in Germany, Netherlands, or France was considered marginally relevant, the information that a website is hosted in Italy, Poland, Spain, Sweden, Belgium, Austria, Ireland, Denmark, Finland, Portugal, or Czechia was considered relevant and the information that a website is hosted in any other EU country was considered important.

The fact that the same SLD might be registered and functioning in other EU country ccTLDs besides .eu was considered relevant since oftentimes individuals, businesses, or institutions decide to offer their content in multiple TLDs in order to increase reach. The fact that these other SLDs might belong to completely different entities discouraged the researchers from considering this information more important.

The ccTLD of the website's registrar was considered of marginal relevance. It might hold a clue as to whether a specific registrar purposefully tries to associate their services with a specific national identity but the registrar's clients, which are the owners of the domains we are investigating, might chose the specific registrar for different reasons.

All ccTLDs collected from a single website under investigation, whether in the homepage or other pages, were tallied. If a specific country's emails were encountered 1 to 5 times, this information was deemed important, while if a specific country's emails were encountered more than 5 times this information became very important. Deciding to include multiple email links to domains using a specific ccTLD different from .eu is a strong indicator of relevancy to that specific country.

The same reasoning was followed for phone numbers collected from tel: links, only in this case even the appearance of 1 to 5 such links was deemed very important. Telephone links not only indicate intention to relate to a specific country but actually confirm physical presence of the entity owning the .eu domain in that specific country.

The relevancy of the information regarding an EU country's name that appeared in full text in English in the homepage or any other sub-page with map links was judged on the basis of the frequency of appearance. Any country that appeared only once was considered marginally relevant, countries that appeared more than once but less than five times were considered relevant and countries that appeared more than five times were considered important. The amount of times a country is repeatedly mentioned in the text of a domain often implies a strong relation with that country.

The information derived from the latitude and longitude of a map identified in a google map link was considered very important since including a map to a location in a specific EU country in a website conveys clear intention to relate to that specific country.

Moving on from collected data regarding an .eu domain's mentioned countries or relevant ccTLDs, the evaluation algorithm focused on utilizing metrics that are based on language detection. In order to convert detected languages into countries of origin, a subroutine was developed that converted a detected language into the corresponding countries. This subroutine was based both on the official languages as indicated by every EU country and on other co-official or minority languages that are spoken in a specific country. These less prevalent languages included Galician, Catalan, Occitan, Basque, Corsican, Breton, Irish, Scottish, Maltese and Kalaallisut.

The information regarding the dominant language appearing in the website of an .eu domain's registrar was considered marginally relevant for similar reasons to the registrar's website ccTLD. This information might hold a hint to the domain's country of origin but it is not definite indication.

Furthermore, the languages that appeared in each domain's homepage were evaluated differently based on if they were English or not. Since the English language is exceptionally dominant in EU websites [35], its presence was not considered an indicator of relation of a website to EU countries where English is an official language, like Ireland, Cyprus, or Malta. In simple terms, if a website's homepage was in English this was considered irrelevant for determining the website's country of origin. If it was any other language, it was considered important information and evaluated as such. The same reasoning was used to evaluate the languages that Common Crawl Index reported as being used in a website and the languages discovered by the Web data extraction algorithm in pages that contained map links.

Despite the very detailed point system and the abundance of collected information, it was impossible to eliminate ties between two or more countries for several domains. A special tie-breaker subroutine was developed that was called upon to decide to which country of origin a website should be attributed to in the case of a tie. This subroutine's algorithm used the population of each individual EU member state as the basis for its decision. The total population of all countries in the tie was calculated with accuracy of 100k people by adding each such country's population. Then a pseudo-number was generated that was used to designate which country won the tie-breaker. Each country had chances to win equal to the percentage of the total population that it represented. Through this methodology any websites that were undecided would be split proportionally between the countries involved. This fact, combined with the very large sample of data, ensured the accuracy of the results. Table 2 presents the various elements that were investigated and the importance attributed to this element by the point system.

**Table 2.** Importance of various examined data.

Data	Importance
Site hosted in country with plenty of data centers	Marginally relevant
Site hosted in country with medium amount of data centers	Relevant
Site hosted in any other country	Important
Country's ccTLD with identical SLD exists	Relevant
Country's ccTLD in the registrar's website	Marginally relevant
Country's coordinates detected in Google map link	Very important
Five or less email links with the country's ccTLD	Important
More than five email links with the country's ccTLD	Very important
Country's telephone code detected in phone link	Very important
Country mentioned once in homepage or map pages	Marginally relevant
Country mentioned two to five times in homepage or map pages	Relevant
Country mentioned more than five times in homepage or map pages	Important
Country corresponds to registrar's website language	Marginally relevant
Country corresponds to homepage language (English)	Marginally relevant
Country corresponds to homepage language (Non-English)	Important
Country corresponds to CC Index language (English)	Marginally relevant
Country corresponds to CC Index language (Non-English)	Important
Country corresponds to map page language (English)	Marginally relevant
Country corresponds to map page language (Non-English)	Important

### 3. Results

As mentioned in the methodology section, the first step of this study was the execution of the .eu domains discovery algorithm which took place in the beginning of April 2022. A total of 203,020 domains were discovered.

The Web data extraction algorithm was executed from mid to late April 2022 on a random sample of 36,395 websites out of the 203,020 collected by the .eu domains discovery algorithm. Out of these, 29,290 were successfully crawled. From the rest, 617 denied access to the crawling algorithm through their robots.txt files and 6488 were unreachable. The total sampling size represents almost 18% of the discovered domains and more than 14% of

successfully accessed domains. As such, it is deemed large enough to offer representative results. The number of domains involved in this research are presented in Table 3.

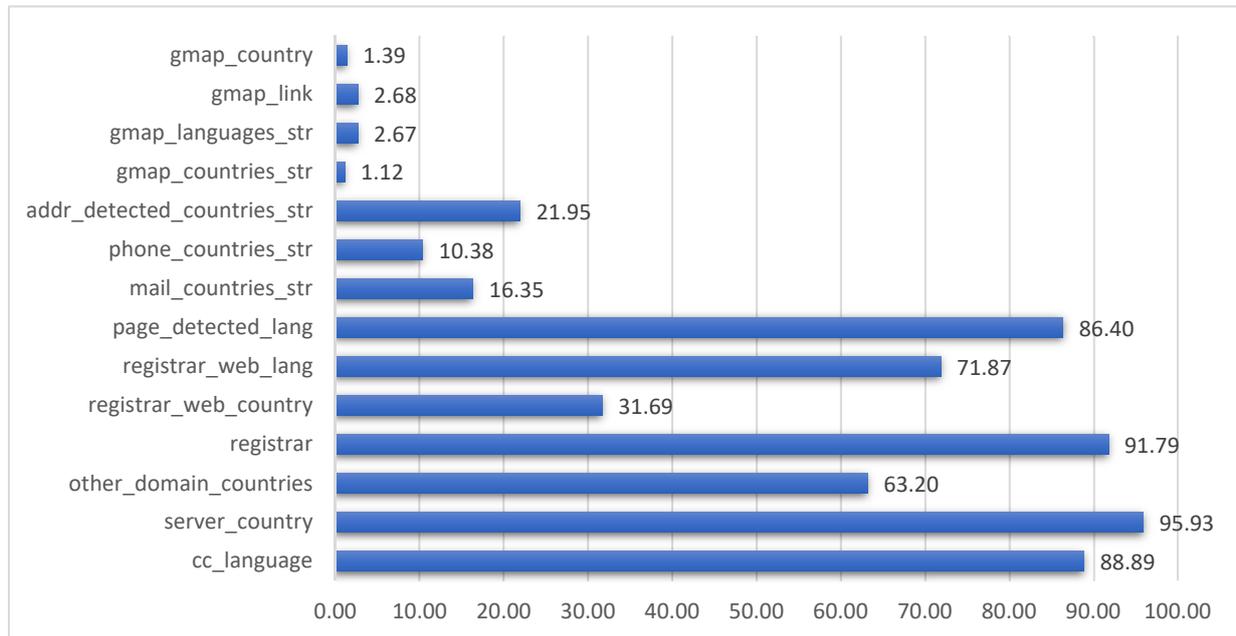
**Table 3.** Number of domains involved in this research.

Metric	Value	Percentage
Discovered domains	203,020	100.00%
Domains randomly sampled	36,395	17.93%
Domains successfully accessed	29,290	14.43%
Unreachable domains	6488	3.20%
Access not allowed by robots.txt	617	0.30%

In addition to discovering .eu domains, an alternative version of the discovery algorithm was used to discover the total amount of pages in the Common Crawl Index for every ccTLD belonging to an EU country. This gives us a metric of how many national domains each country has. By dividing each country’s pages with the total of all EU countries the percentage of total ccTLDs from EU that each national ccTLD represents was derived.

A sample of the information collected by the Web data extraction algorithm is presented in Table 4. The complete information is openly available at a link provided below, in the data availability statement of this article.

As is apparent even from the small sample of Table 4, not all information regarding a domain was always available to be collected by the algorithm. Some were often available, while others seldomly appeared. Table 5 presents every collected variable with the absolute number and the percentage of total domains where it was successfully retrieved. In order to better visualize this information a chart is presented in Figure 5.



**Figure 5.** Chart of percentage of variables’ appearances by successfully crawled domain.

**Table 4.** Sample of data collected by the Web data extraction algorithm.

Id	Domain	CC Lang	Host Countries	SLD Countries	Reg Country	Reg Lang	Home Lang	Mail Country	Phone Country	Text Country	Map Country	Map Lang	Map Country	Reg	Map Link	Date
4	<a href="http://10vkruhu.eu">10vkruhu.eu</a>	slk	SK	sk	sk		sk							<a href="http://www.websupport.sk">www.websupport.sk</a>		14 April 2022 19:30
17	<a href="http://112katowice.eu">112katowice.eu</a>	pol	PL		pl		pl	pl,pl						<a href="http://www.consultingservice.pl">www.consultingservice.pl</a>		16 April 2022 17:52
22	<a href="http://11degrees.eu">11degrees.eu</a>	eng	GB	at,be, de,es, fr,nl,ro			en							<a href="http://domains.meshdigital.com">domains.meshdigital.com</a>		28 April 2022 17:21
32	<a href="http://123-pflege.eu">123-pflege.eu</a>	deu	DE	de	de	de	de							<a href="http://www.vautron.de">www.vautron.de</a>		26 April 2022 10:21
33	<a href="http://1234redes.eu">1234redes.eu</a>	spa	IE		it		es	es,es,es, es,es,es, es,es,es, es,es,es						<a href="http://www.register.it">www.register.it</a>		21 April 2022 7:05
34	<a href="http://123atex.eu">123atex.eu</a>	nld	NL	nl		en	nl		NL,NL, NL,NL, NL,NL, NL,NL					<a href="http://www.openprovider.com">www.openprovider.com</a>		15 April 2022 5:00
39	<a href="http://123consulting.eu">123consulting.eu</a>		LT	at,de,fr, it,lt, nl,uk	lt		en							<a href="http://domains.lt">domains.lt</a>		16 April 2022 9:22
44	<a href="http://123electric.eu">123electric.eu</a>	eng	NL	de,nl		en	en			nl				<a href="http://www.openprovider.com">www.openprovider.com</a>		18 April 2022 22:59
51	<a href="http://123movieshub.eu">123movieshub.eu</a>		US	be,de, it,nl		de								<a href="http://www.key-systems.net">www.key-systems.net</a>		19 April 2022 7:53
62	<a href="http://123tip.eu">123tip.eu</a>	nld,eng		de	nl									<a href="http://www.vimexx.nl">www.vimexx.nl</a>		19 April 2022 14:35

**Table 5.** Variables' appearances by successfully crawled domain.

Variable	Value	Percentage
cc_language	26,037	88.89%
server_country	28,097	95.93%
other_domain_countries	18,510	63.20%
registrar	26,886	91.79%
registrar_web_country	9283	31.69%
registrar_web_lang	21,051	71.87%
page_detected_lang	25,306	86.40%
mail_countries_str	4789	16.35%
phone_countries_str	3041	10.38%
addr_detected_countries_str	6428	21.95%
gmap_countries_str	327	1.12%
gmap_languages_str	783	2.67%
gmap_link	784	2.68%
gmap_country	406	1.39%

The country of origin detection algorithm was executed in early May 2022 and produced results for 28,358 domains out of the 29,290 that were successfully crawled. For the other 932 domains the Web data extraction algorithm didn't produce any useful data that could lead to even a low confidence estimation of the EU country of origin. This includes domains with no data at all and domains with data that point to countries outside the EU as their country of origin. Table 6 presents a sample of data that was produced by the country of origin detection algorithm.

The *country* column presents the final inferred country of origin, the *points* column depicts the number of points it collected from the evaluation system and the column *num* presents the number of collected data that pointed to the selected country of origin.

The columns *Low Points*, *Close Call*, and *Tie* are derivatives of the pointing system. The first one indicates a country selected while another country was at 5 or less points behind, the second one indicates a country selected with a low score of under 20 maximum points, and the third one indicates a country selected while it was tying in first place with one or more different countries.

Table 7 presents how many decisions regarding the country of origin of .eu domains fall under the above categories and what percentage they are of the total of .eu domains.

Out of the 9511 low points, 2582 were attributed to Germany, 1214 to the Netherlands, and 987 to France. Out of 4844 close calls, 2251 were attributed to Germany, 820 to the Netherlands, and 697 to France. Out of the 4353 ties, 1434 were attributed to Germany, 996 to France, 466 to Italy, and 449 to the Netherlands.

In the *Point Details* column, a full account of the points gained is presented, including the country the points were awarded to, the number of points, and the data it was based on. The top scorers column presents all countries within 5 points of the selected country for this domain.

Figure 6 depicts the number of investigated .eu domains that were attributed to each country of origin, as well as the low points, close calls, and attributed ties for each country.

**Table 6.** Sample of data produced by the detection algorithm.

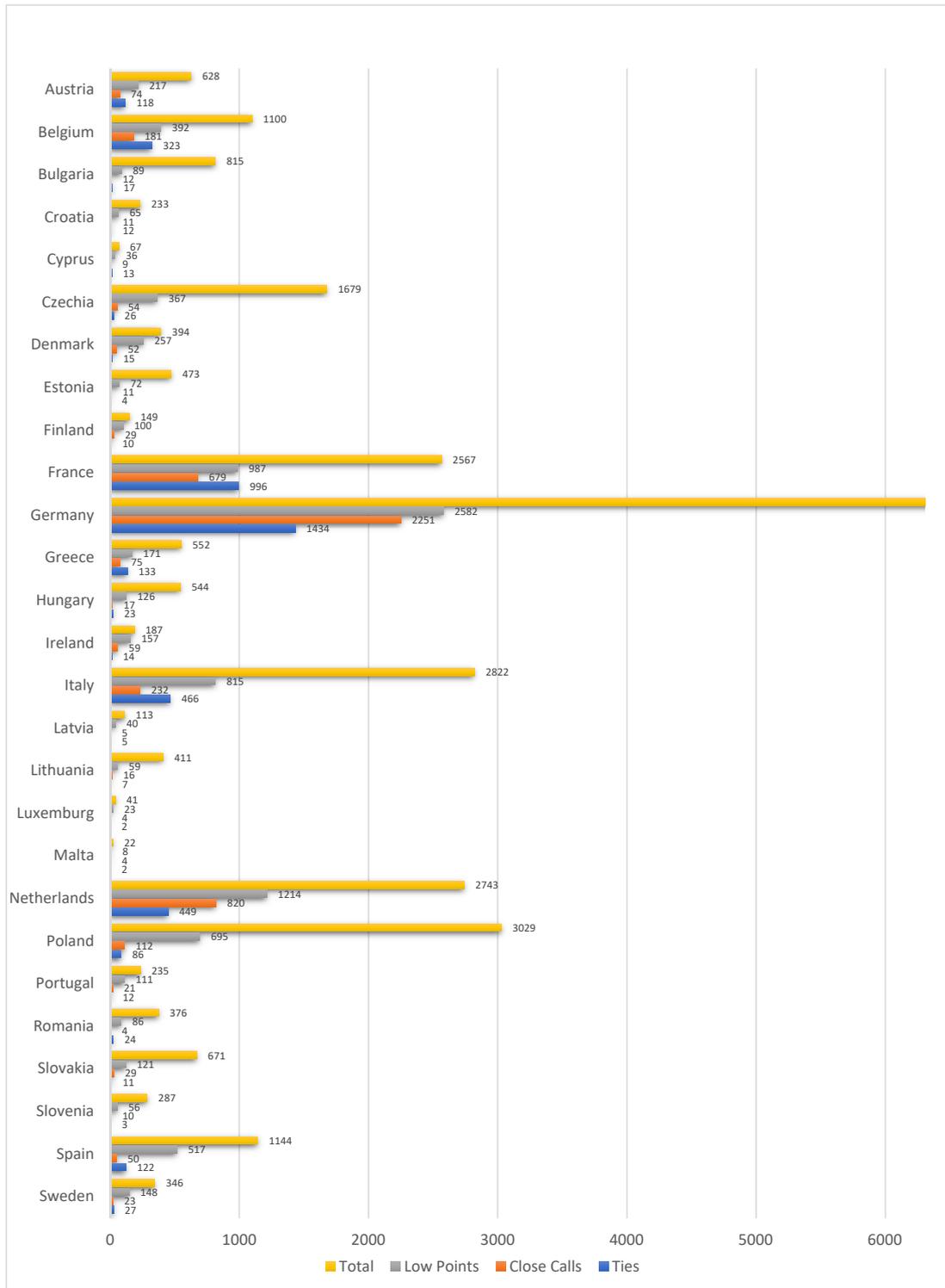
<b>Id</b>	<b>Domain Id</b>	<b>Website</b>	<b>Country</b>	<b>Points</b>	<b>Num</b>	<b>Low Points</b>	<b>Close Call</b>	<b>Tie</b>	<b>Point Details</b>	<b>Top Scorers</b>
1	4	<a href="http://10vkruhu.eu">http://10vkruhu.eu</a>	Slovakia	42	3	0	0	0	Slovakia=>19 (server) Slovakia=>4 (registrant_web) Slovakia=>19 (frontpage_lang)	Slovakia: 42 (3)
2	17	<a href="https://112katowice.eu">https://112katowice.eu</a>	Poland	51	4	0	0	0	Poland=>9 (server) Poland=>4 (registrant_web) Poland=>19 (mails) Poland=>19 (frontpage_lang)	Poland: 51 (4)
3	22	<a href="https://www.11degrees.eu">https://www.11degrees.eu</a>	Germany	9	1	1	0	1	Austria=>9 (other_domains) Belgium=>9 (other_domains) Germany=>9 (other_domains) Spain=>9 (other_domains) France=>9 (other_domains) Netherlands=>9 (other_domains) Romania=>9 (other_domains) United Kingdom=>4 (frontpage_lang) Ireland=>4 (frontpage_lang)	Austria: 9 (1) Belgium: 9 (1) Germany: 9 (1) Spain: 9 (1) France: 9 (1) Netherlands: 9 (1) Romania: 9 (1) United Kingdom: 4 (1) Ireland: 4 (1)
4	32	<a href="https://123-pflege.eu">https://123-pflege.eu</a>	Germany	31	4	0	0	0	Germany=>4 (server) Germany=>4 (registrant_web) Germany=>4 (registrant_web_lang) Austria=>4 (registrant_web_lang) Germany=>19 (frontpage_lang) Austria=>19 (frontpage_lang)	Germany: 31 (4)
5	33	<a href="https://www.1234redes.eu">https://www.1234redes.eu</a>	Spain	48	2	0	0	0	Ireland=>9 (server) Italy=>4 (registrant_web) Spain=>29 (mails) Spain=>19 (frontpage_lang)	Spain: 48 (2)
6	34	<a href="https://123atex.eu">https://123atex.eu</a>	Netherlands	52	3	0	0	0	Netherlands=>4 (server) Netherlands=>29 (phones) United Kingdom=>4 (registrant_web_lang) Ireland=>4 (registrant_web_lang) Netherlands=>19 (frontpage_lang) Belgium=>19 (frontpage_lang)	Netherlands: 52 (3)

Table 6. Cont.

Id	Domain Id	Website	Country	Points	Num	Low Points	Close Call	Tie	Point Details	Top Scorers
7	39	<a href="https://123consulting.eu">https://123consulting.eu</a>	Lithuania	32	3	0	0	0	Lithuania=>19 (server) Austria=>9 (other_domains) Germany=>9 (other_domains) France=>9 (other_domains) Italy=>9 (other_domains) Lithuania=>9 (other_domains) Netherlands=>9 (other_domains) United Kingdom=>9 (other_domains) Lithuania=>4 (registrant_web) United Kingdom=>4 (frontpage_lang) Ireland=>4 (frontpage_lang) Netherlands=>4 (server) Germany=>9 (other_domains) Netherlands=>9 (other_domains) Netherlands=>4 (country_text) United Kingdom=>4 (registrant_web_lang) Ireland=>4 (registrant_web_lang) United Kingdom=>4 (frontpage_lang) Ireland=>4 (frontpage_lang)	Lithuania: 32 (3)
8	44	<a href="https://123electric.eu">https://123electric.eu</a>	Netherlands	17	3	1	0	0	Belgium=>9 (other_domains) Germany=>9 (other_domains) Italy=>9 (other_domains) Netherlands=>9 (other_domains) Germany=>4 (registrant_web_lang) Austria=>4 (registrant_web_lang) Netherlands=>4 (registrant_web) Netherlands=>19 (cc_lang) Belgium=>19 (cc_lang) United Kingdom=>4 (cc_lang) Ireland=>4 (cc_lang)	Netherlands: 17 (3)
9	51	<a href="http://123movieshub.eu">http://123movieshub.eu</a>	Germany	13	2	1	1	0	Belgium=>9 (other_domains) Germany=>9 (other_domains) Italy=>9 (other_domains) Netherlands=>9 (other_domains) Germany=>4 (registrant_web_lang) Austria=>4 (registrant_web_lang) Netherlands=>4 (registrant_web) Netherlands=>19 (cc_lang) Belgium=>19 (cc_lang) United Kingdom=>4 (cc_lang) Ireland=>4 (cc_lang)	Belgium: 9 (1) Germany: 13 (2) Italy: 9 (1) Netherlands: 9 (1)
10	62	<a href="https://www.123tip.eu">https://www.123tip.eu</a>	Netherlands	23	2	0	1	0	Netherlands=>19 (cc_lang) Belgium=>19 (cc_lang) United Kingdom=>4 (cc_lang) Ireland=>4 (cc_lang)	Netherlands: 23 (2) Belgium: 19 (1)

**Table 7.** Domains with .eu ccTLD that were decided.

Column	Number of Domains	Percentage
Low Points	9511	33.54%
Close Call	4844	17.08%
Tie	4353	15.35%



**Figure 6.** Domains with the .eu ccTLD by country out of a sample of 29,091 domains.

#### 4. Discussion

With a first glance at the metrics of Table 3 it becomes apparent that the total number of .eu domains investigated in this research is fairly representative. The sum of .eu TLD websites as provided by ccIndex were recorded and the sample that was selected for data extraction had the same probability of selection and came from the total collected websites without any bias regarding website size, availability, SLD naming, or other factors. The fact that the total domains successfully crawled were more than 14.4% of the total websites gathered, combined with the sampling method, provided us with a varied and robust data sample.

The various variables that were recorded by the Web data extraction algorithm as presented in Table 5, with the addition of the *cc\_language* metric that was provided by Common Crawl Index during the domain discovery process, paint an overall picture of each domain investigated. As is made abundantly clear, not all variables are equally represented in each domain. Three major brackets of recorded data emerge:

- Bracket A: Data that can be found for the majority of domains (50%+) which include *cc\_language*, *server\_country*, *other\_domain\_countries*, *registrar*, *registrar\_web\_lang* and *page\_detected\_lang*.
- Bracket B: Data that can be found in a smaller but still significant number of domains (10–50%) which include *registrar\_web\_country*, *mail\_countries\_str*, *phone\_countries\_str* and *addr\_detected\_countries\_str*.
- Bracket C: Data that can be found in very few domains (<10%) which include *gmap\_countries\_str*, *gmap\_languages\_str*, *gmap\_link*, *gmap\_country*.

Studying the recorded data indicated that ~58% of domains investigated involved at least one metric from brackets B and C. This means that the majority of country of origin attributions by the algorithm were based on data from multiple brackets. Bracket C that revolved around the identification and parsing of map links was only present in 2.68% of the domains, but since all of this bracket's variables were considered important or very important, it still played an active role in producing the final results.

Figure 6 presents the final distribution of the sample domains in each country of origin alongside the ties, close calls and low confidence attributions, creating a very interesting picture of the landscape of .eu domains.

A large number of ties per total domains was identified in France, Belgium, Greece, and Germany (>20% of their total domains were the result of ties). Sharing a language and geographical proximity seem to be a key factor to generating ties. Cyprus, Austria, Italy, Netherlands, and Malta also follow with a high percentages of ties (>9%) amplifying the connection between algorithm ties and linguistic ties.

Almost the same group of countries seem to rank high in close calls as a percentage of total domains too, with Germany, Ireland, France, and the Netherlands over 25%, followed by Finland, Malta, Belgium, Greece, Cyprus, Denmark, and Austria over 10%. The same factors that lead to ties can be attributed to leading to close calls. Linguistic similarities, geographical proximity, and cultural or business relations all play a role in creating a situation where it would be hard to attribute a website to one specific country of origin even after a complete human-performed audit. The countries with many ties and close calls form cultural and linguistic pairs or subgroups, where detecting the country of origin of a website can be harder (e.g., Greece–Cyprus, Belgium–France, Belgium–Netherlands, Germany–Austria, etc.).

An outlier in the case of close calls is Ireland, which seems to have a much larger percentage than expected. The fact that the English language was not attributed any points probably played a large part in this case. Continuing with the same trend, from the domains that were attributed despite a low point count, most belong to Ireland with more than 80% of its attributions having low points. This is also a side-effect of the reduced point score of detecting the English language in either the homepage or other pages of a domain. Ireland is generally an outlier in almost all of the above observations. This is arguably the result of the decision to not consider the use of the English language an indicator of country

of origin. Unfortunately, any other form of scoring for English leads to vast numbers of dubious .eu domains being attributed to Ireland solely for language reasons, which creates an even bigger outlier.

Finland and Denmark also present a high percentage of low point domain attributions, with Luxembourg and Cyprus following close behind to form a group of countries with >50% low point attributions.

In order to gain a first idea of the relation between .eu usage per country and its economic power, a comparison can be made using each country’s GDP as a percentage of the total GDP of the European Union. The GDP of each country was based on official EU statistics as provided by Eurostat [36]. Figure 7 presents the percentages of each country’s GDP out of the total EU GDP, alongside the percentages of each country’s .eu domains out of the total .eu domains investigated. Looking at countries that represent more than 3% of the total EU GDP, it appears that Poland and the Netherlands have much larger representation in the .eu domain than their GDP would suggest. On the other hand, France, Italy, and Spain hold a much smaller presence, while Germany almost strikes a balance. The same irregularity continues in smaller countries, with very few displaying a balance between the two metrics.

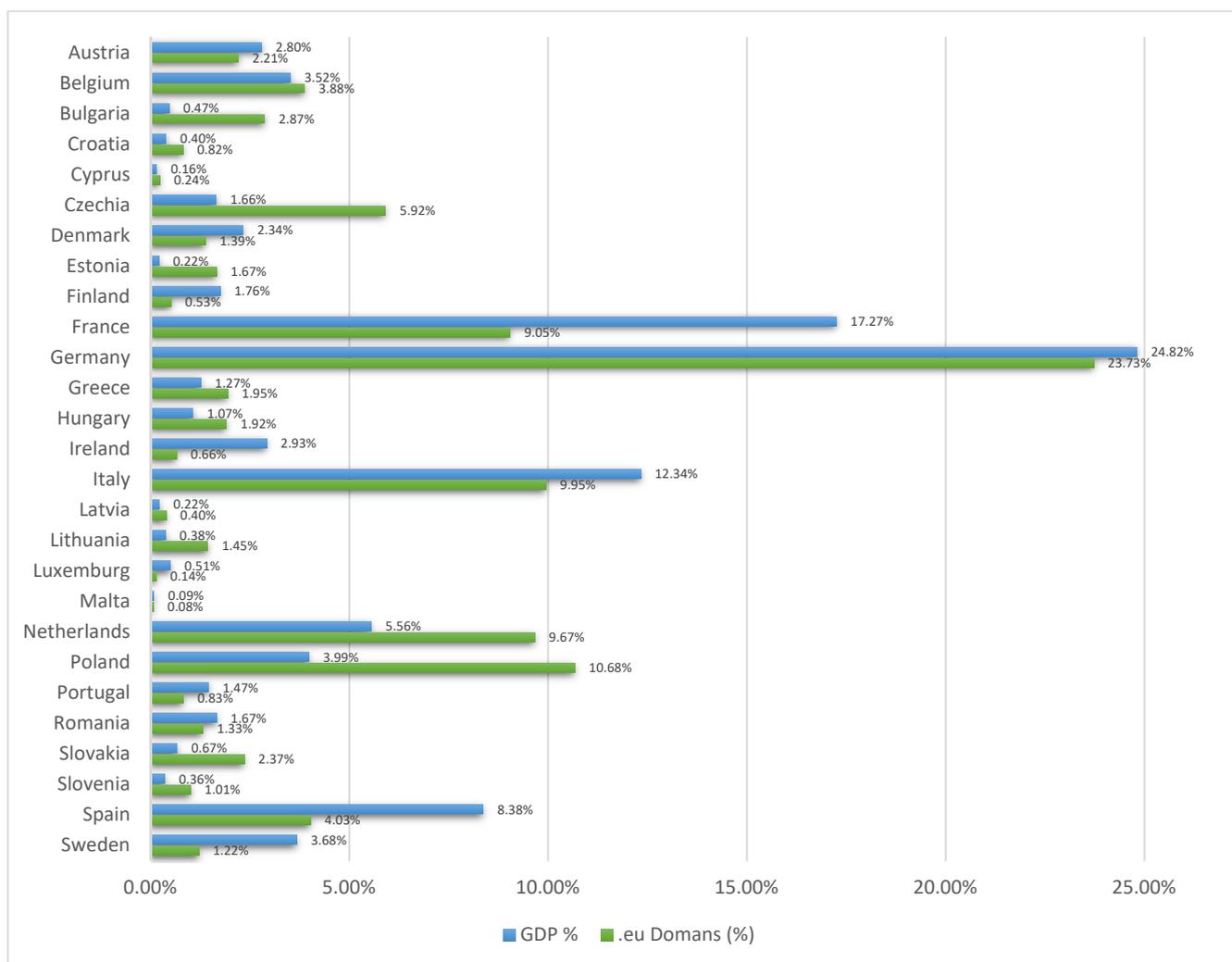


Figure 7. Percentage of EU domains vs. percentage of GDP per country.

In order to investigate whether this is indicative of a larger Web footprint for these specific countries, or a preference towards the European Internet identity that is inherent

in the .eu usage, a comparison must be made between the national ccTLDs coverage as collected by the ccIndex and GDP. Figure 8 presents the percentage of discovered domains belonging to national ccTLDs out of the sum of all national ccTLDs belonging to EU countries, alongside the GDP percentages. Investigating the countries with GDP over 3% of the total GDP of the EU, we notice that Germany, Italy, Poland, and the Netherlands display a decrease in domain coverage in national domains as opposed to .eu domains. At the same time France, Romania, and Spain display an increase.

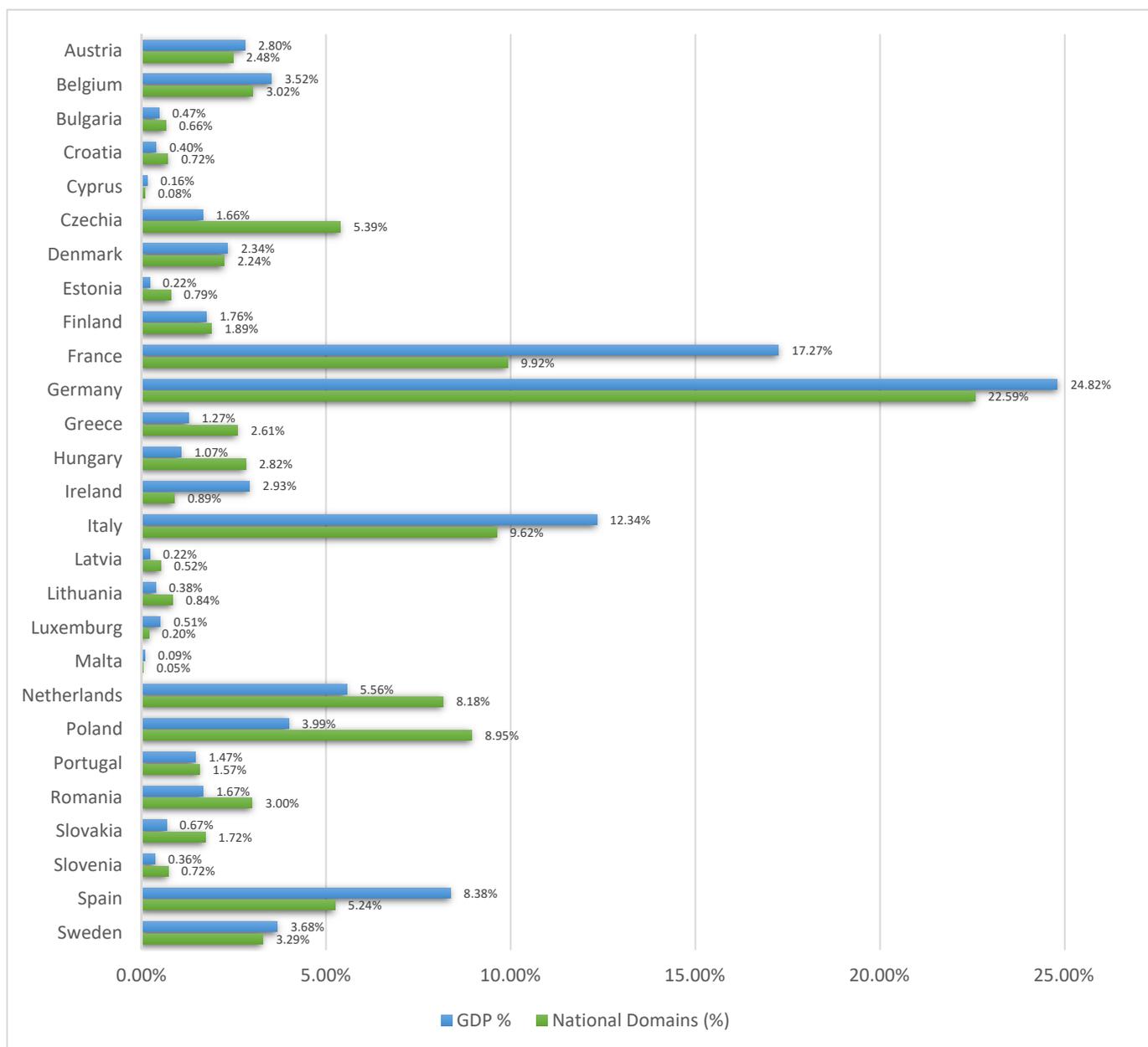


Figure 8. Percentage of national domains vs percentage of GDP per country.

The above charts in Figures 7 and 8 give us a clear overview of both .eu and national ccTLD domain distribution, but in order to draw conclusions regarding each country’s endeavor to make use of the .eu TLD it is important to disconnect the metric from the size of the country’s economy. Dividing the percentage of .eu or national domains with the percentage of GDP gives us two ratios, ( $eu/gdp$  and  $nat/gdp$ ), that indicate the relation between a country’s presence in each ccTLD landscape and its GDP. If the ratio is higher

than 1 it implies a larger footprint in the equivalent ccTLD landscape. Table 8 presents these ratios for each country.

**Table 8.** Ratios referring to each country’s relevant footprint in the domain usage landscape.

Country	<i>eu/gdp</i>	<i>nat/gdp</i>
Austria	0.79	0.88
Belgium	1.10	0.86
Bulgaria	6.08	1.39
Croatia	2.07	1.81
Cyprus	1.48	0.52
Czechia	3.56	3.24
Denmark	0.59	0.96
Estonia	7.74	3.65
Finland	0.30	1.08
France	0.52	0.57
Germany	0.96	0.91
Greece	1.53	2.05
Hungary	1.79	2.64
Ireland	0.23	0.30
Italy	0.81	0.78
Latvia	1.79	2.34
Lithuania	3.79	2.20
Luxembourg	0.28	0.39
Malta	0.86	0.54
Netherlands	1.74	1.47
Poland	2.68	2.24
Portugal	0.57	1.07
Romania	0.79	1.80
Slovakia	3.51	2.56
Slovenia	2.80	2.00
Spain	0.48	0.63
Sweden	0.33	0.89

Figure 9 presents the *eu/gdp* ratio over a map of Europe while Figure 10 presents the *nat/gdp* ratio over a map of Europe. Visible at first glance is the differentiation between Western and Eastern Europe, which is to a large extent a result of the “clear economic underpinnings of the East–West divide” [37]. This is an observation more connected to the economic disparity of the east than to its presence on the .eu TLD. Even so, comparing the two maps can shed light on some differences. Northern countries such as Sweden, Finland, and Denmark all present a significant smaller .eu presence compared to their ccTLD presence. This could be a remnant of Nordic hesitance towards the EU as it was documented during the years of their accession [38] and is noticeable even in today’s trends [26].

In order to investigate the relationship between these two ratios, the Spearman’s Rho correlation coefficient was calculated and the results are presented in Table 9. A positive correlation between the two ratios was detected, that was significant at the 0.00 level (2-tailed). This means that the null hypothesis, that there is no monotonic relationship between the two variables, is rejected. It is safe to assume that the national domains usage per GDP ratio, which was measured through discovering the available national ccTLD domains in Common Crawl Index, has a significant correlation with the ratio of .eu domain usage per GDP, which was derived by the country detection algorithm.

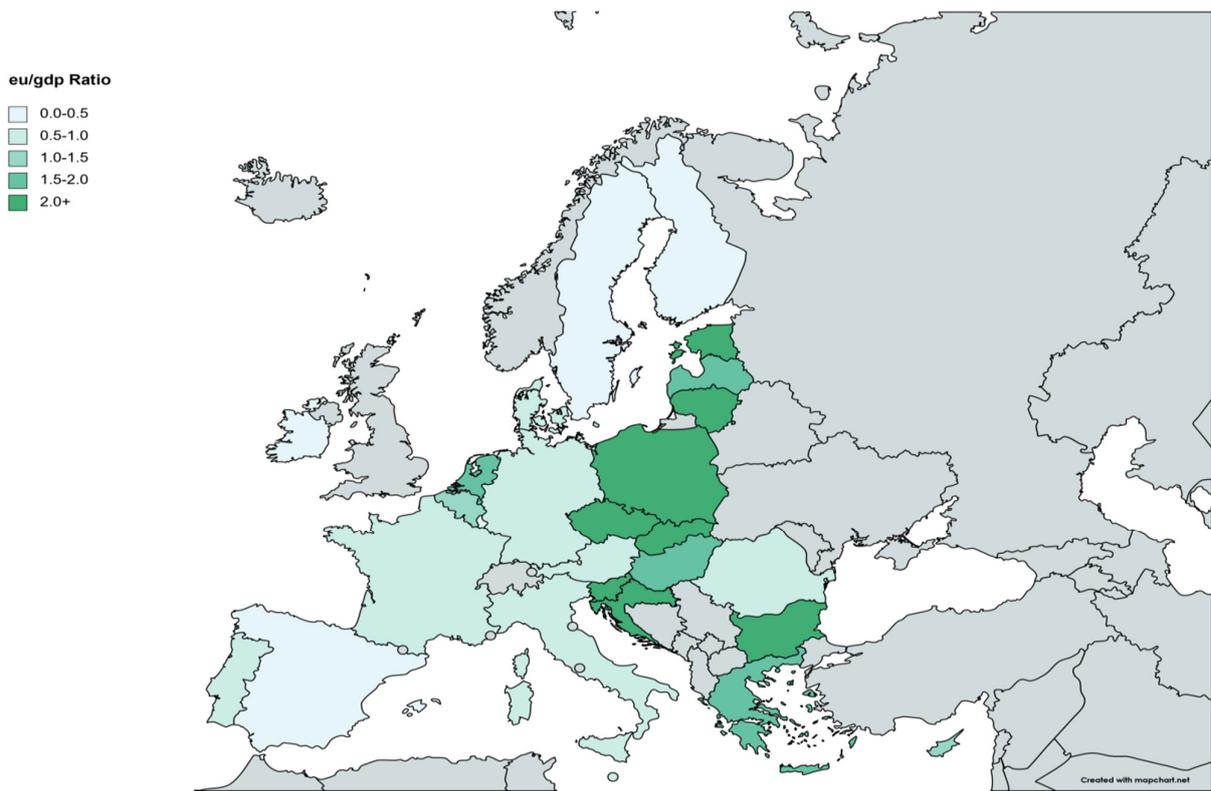


Figure 9. *eu/gdp* ratio by EU member State.

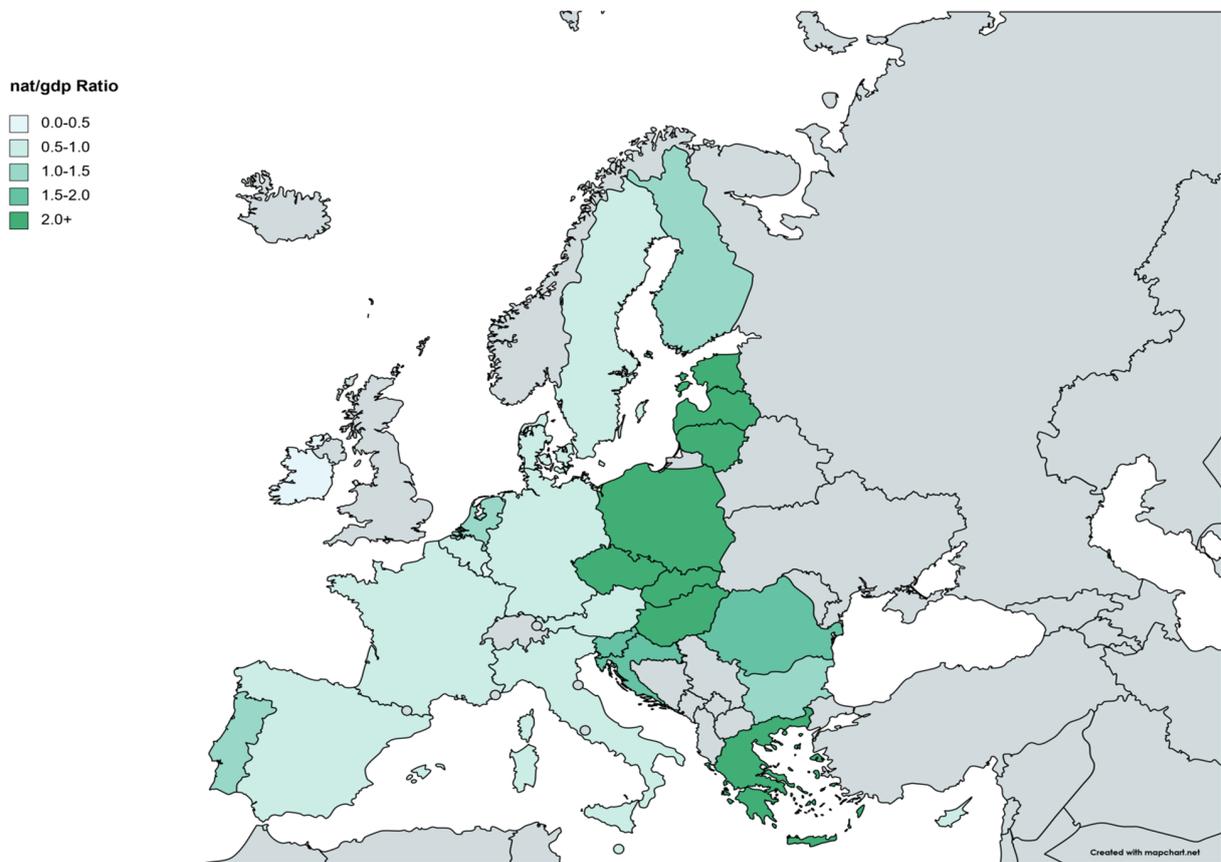


Figure 10. *nat/gdp* by EU member state.

**Table 9.** Spearman's correlation coefficient between the ratios *eu/gdp* and *nat/gdp*.

	$r_s$	$p$ (2-Tailed)	N
<i>eu/gdp</i> <i>nat/gdp</i>	0.7658	0.00	27

This correlation is a strong indicator of the quality of the country detection algorithm, because it indicates that each country's presence in the .eu domain is closely related but not identical with each country's overall World Wide Web presence as measured by the usage of its national ccTLD. Being closely related indicates a general trend in some countries to have a more active presence on the Internet than others. Not being overly correlated indicates that there are other significant factors at play when the choice to adopt the .eu TLD is made, apart from this general trend.

As detailed in the introduction, one of the main roles for creating the .eu TLD was to contribute to the Digital Market by providing EU residents and businesses a common Web domain that inspires confidence [7], stability, and security. In order to investigate further whether that purpose is being fulfilled, trade statistics from the EU [39] were used to obtain the ratio of EU external trades (Extra-EU) by GDP for each country. Table 10 presents the collected numerical data for external trade, the data as a percentage of total EU external trade, and the ratio of that external trade percentage divided by each country's GDP as a percent of total EU GDP (variable *ex/gdp*).

**Table 10.** EU external trade statistics (Extra-EU) and derived data.

Country	Extra-EU (Values in 1 m €)	Extra-EU (%)	<i>ex/gdp</i>
Austria	57,579	2.34%	0.83
Belgium	181,941	7.38%	2.10
Bulgaria	12,195	0.49%	1.05
Croatia	6244	0.25%	0.64
Cyprus	2560	0.10%	0.65
Czechia	44,750	1.82%	1.09
Denmark	55,751	2.26%	0.97
Estonia	6640	0.27%	1.25
Finland	33,122	1.34%	0.76
France	252,862	10.26%	0.59
Germany	699,363	28.38%	1.14
Greece	19,782	0.80%	0.63
Hungary	29,295	1.19%	1.11
Ireland	117,242	4.76%	1.63
Italy	269,906	10.95%	0.89
Latvia	7701	0.31%	1.41
Lithuania	15,987	0.65%	1.70
Luxembourg	3063	0.12%	0.24
Malta	1250	0.05%	0.56
Netherlands	256,438	10.41%	1.87
Poland	86,698	3.52%	0.88
Portugal	21,431	0.87%	0.59
Romania	21,924	0.89%	0.53
Slovakia	20,297	0.82%	1.22
Slovenia	16,358	0.66%	1.84
Spain	140,367	5.70%	0.68
Sweden	83,265	3.38%	0.92

The ratio calculated by dividing the external trade percentage by the GDP of a country provides us a metric of how competitive and outward-looking that country's economy is in terms of exports to countries outside of the EU, while at the same time diminishing the influence of the size of that country's economy as determined by GDP.

In order to investigate whether a country's presence in the .eu TLD was measured by the *eu/gdp* ratio and its presence in EU's national ccTLDs as measured by the *nat/gdp* ratio, its *eu/gdp* ratio Spearman's rho correlation coefficient was calculated for both cases and is presented in Table 11.

**Table 11.** Spearman's Rho correlation coefficient between the ratios *eu/gdp* and *ex/gdp*, and *nat/gdp* and *ex/gdp*.

	$r_s$	$p$ (2-Tailed)	N
<i>eu/gdp</i> <i>ex/gdp</i>	0.4648	0.01458	27
<i>nat/gdp</i> <i>ex/gdp</i>	0.36819	0.05881	27

The association between the two variables *eu/gdp* and *ex/gdp* can be considered statistically significant by normal standards. This leads to the rejection of the null hypothesis (that there is no correlation between the variables), confirming their positive monotonic relationship. This means that when observing, an increase in the *eu/gdp* variable will oftentimes be followed by an increase in the *ex/gdp* variable.

The association between the two variables *nat/gdp* and *ex/gdp* cannot be considered statistically significant by normal standards, although it is relevantly close to significance. This leads to the acceptance of the null hypothesis, that there is no correlation between the variables.

The above measurements, by all means, indicate that each country's use of the .eu TLD is connected to its external trade in a way that is independent from its economy's size as measured by its GDP. In simpler terms, countries that choose to invest and expand in the landscape of the .eu TLD are more likely to have a higher external trade by GDP ratio. They are more likely to have an economy that benefits from exports to countries outside the EU. From an e-commerce perspective, countries that have or are aspiring to have better outreach beyond EU borders are more inclined to present themselves as part of the Union in order to capitalize on the EU's reputation for stability and security.

Based on the above consideration, the interrelation between *eu/gdp* and *ex/gdp* was again investigated, this time for the 10 EU countries with the lowest external trade revenues in millions of euros. Based on Table 10 these countries are Malta, Cyprus, Luxembourg, Croatia, Estonia, Latvia, Bulgaria, Lithuania, Slovenia, and Greece, all with external trade below 20 billion €. Table 12 presents the results of calculating Spearman's Rho coefficient for these countries.

**Table 12.** Spearman's Rho correlation coefficient between the ratios *eu/gdp* and *ex/gdp* for N = 10 with the countries with the lowest external trade.

	$r_s$	$p$ (2-Tailed)	N
<i>eu/gdp</i> <i>ex/gdp</i>	0.72121	0.01857	10

The null hypothesis that there is no correlation between the two variables for this specific data-set can be rejected. Moreover, the value of Rho indicates that countries with smaller total revenue from external trade display an even stronger correlation between their .eu usage by GDP ratio and their external revenue by GDP ratio. This measurement is a strong indicator that for countries with low established external trade, increase of the usage of the .eu domain is a driving force behind increase of the total external trade revenue. In fact, there have been initiatives in the past to help such countries increase trading through internet use (for example in Malta [40]) so it is not surprising that the .eu TLD is another tool to that end. Similar findings regarding correlation between external trade and internet penetration also appear in Middle East and North African countries [41].

## 5. Conclusions

In this study, an automated system based on Web data extraction and evaluation algorithms was presented. This system collected a large sample of .eu websites and inferred their original country of origin, making use of a series of metrics regarding the website's various technical and content related aspects. Through its use, the landscape of the .eu TLD usage by each country was outlined.

By defining the *eu/gdp* ratio of that usage divided by the country's GDP as a percentage of the total GDP of the European Union, the involvement of each country in the .eu TLD was measured independently from its economic size. The interrelation of this ratio and the ratio of each country's national ccTLD usage divided by its GDP percentage (*nat/gdp*) was investigated and the conclusion was that there is a strong significant correlation between them, which acts as a confirmation to an extent of the ability of the algorithm to infer each .eu website's country of origin.

By investigating the interrelation between both these ratios and the *ex/gdp* ratio that presents the importance of a country's external trade in relevance with its economic size, it was concluded that involvement in the .eu TLD and external trade are significantly correlated even when the country's economic size is removed from the picture. Especially when considering countries with lower total external trade revenue, the relationship between the *eu/gdp* and *ex/gdp* variables is even stronger.

The main limitations of this research appeared in the process of detecting the country of origin. Some notable weak points have already been pointed out, such as the difficulty of inferring Ireland's presence because of the common use of the English language, or the inability to attribute a country of origin to all such domains. For some domains the concept of country of origin might not necessarily apply due to their international nature. Despite these, the algorithm put a lot of effort into handling edge cases and the result was satisfactory. Another limitation was the use of the Common Crawl Index to identify both the .eu domains and to determine the landscape of national ccTLDs. Although ccIndex data is almost certainly representative, there may be value in repeating the process with websites selected not randomly, but based on their popularity as derived through data traffic from a digital intelligence provider. Finally, it should be emphasized that the Spearman correlation between *eu/gdp* and *ex/gdp*, as any Spearman correlation indicates a monotonic relationship only between the input variables and not a linear relationship, and it does not establish causality. There was no proportionality discovered in the increases of each variable.

Using a similar methodology in the future, other TLDs can be investigated for their websites' countries of origin beyond the scope of the EU. Analyzing global TLD usage worldwide, especially outside the US, penetration of newer gTLDs in different regional markets and other similar issues can lead to interesting conclusions about the connection of TLDs with economic, social, or even political indexes. Alternatively, a similar methodology can be used to measure the influence of larger countries within their sphere of economic or geopolitical influence. For example, identifying USA oriented websites in Central and South America, or China oriented websites in East and South East Asia. Finally, it might be a good idea to study the relationships established in this article across a broader time period, either by regularly repeating the study and collecting data in the future, or through usage of archival website data as they can be retrieved from internet archival initiatives.

The World Wide Web is the perfect means to reach audience beyond a country's own borders both culturally and of course commercially. But in order to achieve results, recognition and confidence are required. The .eu TLD's relation with an EU member state's external trade revenue is a strong indicator that that recognition and confidence may be found in the unified European Internet identity that it was created to embody, especially for member states that do not have an already established international commercial presence.

**Author Contributions:** Conceptualization, A.G.; Data curation, A.G., M.P. and A.K.; Formal analysis, M.P. and A.K.; Investigation, A.G., M.P. and L.L.; Methodology, A.G., M.P. and L.L.; Project administration, A.G. and M.P.; Resources, L.L.; Software, M.P.; Supervision, A.G.; Validation, A.G., M.P. and

A.K.; Visualization, M.P. and A.K.; Writing—original draft, M.P. and L.L.; Writing—review & editing, A.G., M.P. and L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in Zenodo at [<https://doi.org/10.5281/zenodo.6564241>], (accessed on 19 May 2022) reference number [10.5281/zenodo.6564241]. The source code of the algorithms presented in this study are openly available in Zenodo at (<https://doi.org/10.5281/zenodo.6604080>), (accessed on 1 June 2022) reference number (10.5281/zenodo.6604080).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Ershov, Y.M. National identity in new media. *Procedia—Soc. Behav. Sci.* **2015**, *200*, 206–209. [[CrossRef](#)]
- Vrysis, L.; Vryzas, N.; Kotsakis, R.; Saridou, T.; Matsiola, M.; Veglis, A.; Arcila-Calderón, C.; Dimoulas, C. A Web Interface for Analyzing Hate Speech. *Future Internet* **2021**, *13*, 80. [[CrossRef](#)]
- IANA. Root Zone Database. Internet Assigned Numbers Authority IANA. Available online: <https://www.iana.org/domains/root/db> (accessed on 15 May 2022).
- Wahdani, F.; Alfaouri, M. .eu Top Level Domain Name & Free Movement of Services: The Eu Policy over Single Digital Market. *Cross Cult. Manag. J.* **2020**, *1*, 53–66.
- Van Gelder, S. Dot EU—The first decade. EURid Libr. Belg. 2016. Available online: [https://eurid.eu/media/filer\\_public/d3/85/d38538c1-dac5-4e28-a779-cc31b8259697/boek\\_dot\\_eu\\_v05.pdf](https://eurid.eu/media/filer_public/d3/85/d38538c1-dac5-4e28-a779-cc31b8259697/boek_dot_eu_v05.pdf) (accessed on 15 May 2022).
- European Parliament. Regulation (EC) No 733/2002 of the European Parliament and of the Council of 22 April 2002 on the implementation of the.eu Top Level Domain. *Off. J.* **2002**, *113*, 1–5. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32002R0733> (accessed on 15 May 2022).
- Shaping Europe’s Digital Future. The Top-Level Domain.eu. Available online: <https://digital-strategy.ec.europa.eu/en/policies/eu-top-level-domain> (accessed on 15 May 2022).
- Wass Schlesinger, E. *Addressing the World: National Identity and Internet Country Code Domains*; Rowman & Littlefield: Lanham, MD, USA, 2003; 186p. Available online: <http://www.loc.gov/catdir/toc/ecip043/2003009131.html> (accessed on 10 May 2022).
- Ørmen, J.; Helles, R.; Bruhn Jensen, K. Converging cultures of communication: A comparative study of Internet use in China, Europe, and the United States. *New Media Soc.* **2021**, *23*, 1751–1772. [[CrossRef](#)]
- Webnic. 10 Reasons Why You Should Register a.eu Domain. 2020. Available online: <https://www.webnic.cc/10-reasons-why-you-should-register-a-eu-domain/> (accessed on 31 May 2022).
- Shaping Europe’s digital future. 10 years of.eu! *Digibyte*. 2016. Available online: <https://digital-strategy.ec.europa.eu/en/news/10-years-eu> (accessed on 15 May 2022).
- Nolan, P.; McMahon, R. EverCloserUnion. *eu. Comput. Law Rev. Int.* **2006**, *7*, 17–21. [[CrossRef](#)]
- EURid. EURid and the Ecommerce Foundation Strengthen Their Collaboration. 2020. Available online: <https://eurid.eu/en/news/eurid-ecommerce-collaboration> (accessed on 15 May 2022).
- Sassatelli, M. Imagined Europe: The shaping of a European cultural identity through EU cultural policy. *Eur. J. Soc. Theory* **2002**, *5*, 435–451. [[CrossRef](#)]
- Aichholzer, J.; Kritzinger, S.; Plescia, C. National identity profiles and support for the European Union. *Eur. Union Polit.* **2021**, *22*, 293–315. [[CrossRef](#)]
- Smith, A.D. National identity and the idea of European unity. *Int. Aff.* **1992**, *68*, 55–76. [[CrossRef](#)]
- Carey, S. Undivided loyalties: Is national identity an obstacle to European integration? *Eur. Union Polit.* **2002**, *3*, 387–413. [[CrossRef](#)]
- Anderson, B. *Imagined Communities: Reflections on the Origins and Spread of Nationalism*; Verso: London, UK, 1983.
- Wodak, R.; de Cillia, R.; Reisigl, M.; Liebhart, K.; Hirsch, A.; Mitten, R.; Unger, J.W. *The Discursive Construction of National Identity*; Edinburgh University Press: Edinburgh, UK, 2009. Available online: <https://www.jstor.org/stable/10.3366/j.ctt1r26kb> (accessed on 10 May 2022).
- Cram, L. Does the EU need a navel? Implicit and explicit identification with the European Union. *JCMS J. Common Mark. Stud.* **2012**, *50*, 71–86. [[CrossRef](#)]
- Bruter, M. Winning hearts and minds for Europe: The impact of news and symbols on civic and cultural European identity. *Comp. Polit. Stud.* **2003**, *36*, 1148–1179. [[CrossRef](#)]
- Fligstein, N. The process of europeanization. *Polit. Eur.* **2000**, *1*, 25–42. [[CrossRef](#)]

23. Ariely, G. Globalization, immigration and national identity: How the level of globalization affects the relations between nationalism, constructive patriotism and attitudes toward immigrants? *Group Process. Intergroup Relat.* **2012**, *15*, 539–557. [CrossRef]
24. Pew Research Center. *European Public Opinion Three Decades after the Fall of Communism*; Pew Research Center: Washington, DC, USA, 2019.
25. Pelikánová, R. And the best top level domain for European Enterprises is . . . . *Int. Comp. Law Rev.* **2012**, *12*, 43–59. [CrossRef]
26. European Commission; Directorate-General for Communication; Joint Research Centre. *Values and identities of EU Citizens: Report*; Publications Office: Luxembourg, 2021. [CrossRef]
27. EURid. Brexit Notice. 2022. Available online: <https://eurid.eu/en/register-a-eu-domain/brexit-notice/> (accessed on 4 May 2022).
28. Common Crawl. Common Crawl—In a Nutshell, Here’s Who We Are. Available online: <https://commoncrawl.org/about/> (accessed on 31 May 2022).
29. MaxMind. GeoLite2 Free Geolocation Data. Available online: <https://dev.maxmind.com/geoip/geolite2-free-geolocation-data?lang=en> (accessed on 31 May 2022).
30. MaxMind. GeoIP2 PHP API. Available online: <https://maxmind.github.io/GeoIP2-php/> (accessed on 31 May 2022).
31. Shur, P. Language-Detection. Available online: <https://github.com/patrickschur/language-detection> (accessed on 31 May 2022).
32. Giggs, J. Libphonenumber for PHP. Available online: <https://github.com/giggsey/libphonenumber-for-php> (accessed on 31 May 2022).
33. Nominatim. Nominatim—Introduction. Available online: <https://nominatim.org/release-docs/latest/> (accessed on 31 May 2022).
34. Statista. Number of Data Centers in Europe by Country. 2021. Available online: <https://www.statista.com/statistics/878621/european-data-centers-by-country/> (accessed on 4 May 2022).
35. Giannakouloupoulos, A.; Pergantis, M.; Konstantinou, N.; Lamprogeorgos, A.; Limniati, L.; Varlamis, I. Exploring the Dominance of the English Language on the Websites of EU Countries. *Future Internet* **2020**, *12*, 76. [CrossRef]
36. EC Europa. Eurostat Statistics—GDP and Main Components (Output, Expenditure and Income). Available online: [https://ec.europa.eu/eurostat/databrowser/view/NAMA\\_10\\_GDP/default/table?lang=en&category=na10.nama10.nama\\_10\\_ma](https://ec.europa.eu/eurostat/databrowser/view/NAMA_10_GDP/default/table?lang=en&category=na10.nama10.nama_10_ma) (accessed on 4 May 2022).
37. Volintiru, C.; Bargaoanu, A.; Stefan, G.; Durach, F. East-West Divide in the European Union: Legacy or Developmental Failure? *Rom. J. Eur. Aff.* **2021**, *21*, 93–118.
38. Torsten, S.; Sanneke, K. Shared hesitance, joint success: Denmark, Finland, and Sweden in the European Union policy process. *J. Eur. Public Policy* **2005**, *12*, 157–176. [CrossRef]
39. AllThatStatsNow. Intra-Extra-EU Trade Statistics. Available online: <https://now.allthatstats.com/cntrade/cn0xxxxxx-0-011988-total-trade> (accessed on 4 May 2022).
40. Said, A. Helping small firms trade effectively with the Internet. *Int. Trade Forum* **2000**, *3*, 16–19.
41. Zhang, X. Digital Divide and External Trade Liberalization in the MENA Region: A Theoretical and Empirical Investigations. In *Key Challenges and Policy Reforms in the MENA Region. Perspectives on Development in the Middle East and North Africa (MENA) Region*; Ben Ali, M.S., Ed.; Springer: Cham, Switzerland, 2022; pp. 103–121. [CrossRef]