International Journal of Web Applications

DLINE JOURNALS

Print ISSN: 0974-7710
Online ISSN: 0974-7729

IJWA 2026: 18 (3)
https://doi.org/10.6025/ijwa/2026/18/3/125-144

Comprehensive Analysis of DNS Reliability and Web Reachability in the German.de Domain Inventory

Ramiro Sámano Robles
Instituto Superior de Engenharia
do Porto, Rua
Porto
Portugal
RASRO@isep.ipp.pt

ABSTRACT

The Domain Name System (DNS) is a critical foundation of internet functionality, yet its reliability remains vulnerable to physical, logical, and operational failures. While substantial research exists on DNS security and topology, large scale assessments of country code top level domains (ccTLDs) remain limited. This study presents a comprehensive empirical analysis of DNS reliability and web reachability within the German .de domain ecosystem. Utilizing a stratified 10% random sample (N = 636,853) from a verified inventory of 6.38 million. de domains (derived from Common Crawl data spanning 2019–2026), we evaluated sequential DNS resolution and HTTP/HTTPS accessibility. Descriptive and inferential statistical methods, along with a novel DNS Reliability Index (DRI), were used to quantify operational health. Results indicate a DNS resolution rate of 66.24% (95% CI [66.12%, 66.37%]), but a significant attrition to full web reachability, with only 35.42% of the total sample (or ~42.2% of actively checked domains) returning successful HTTP 200 responses. DNS failures (ERR_DNS, 28.36%) and network timeouts (ERR_TIMEOUT, 11.94%) emerged as the primary barriers to accessibility. The DRI remained stable (~66.2%) across domain name lengths, though temporal anomalies highlighted batch specific probing effects. These findings Shown that domain registration does not guarantee service availability, highlighting the need for continuous monitoring and improved registration maintenance to enhance the resilience of the web ecosystem.

Keywords: DNS Reliability, Web Reachability, German.de Domain, Internet Resilience, DNS Resolution, HTTP Probing, DNS Reliability Index (DRI), Domain Attrition.

Received: 18 December 2025, Revised: 20 April 2026, Accepted: 3 June 2026

1. Introduction

The Domain Name System (DNS) is a fundamental component of the Internet that translates human readable domain names into IP addresses, enabling users and applications to access online services efficiently [1]. Without DNS, websites and Internet based applications would be unable to locate the servers hosting the requested resources, thereby disrupting normal Internet functionality. Owing to its critical role, DNS has been designed as a distributed and redundant system in which DNS records are replicated across multiple name servers, preferably located on different networks, to ensure service continuity during network failures or outages [2].

Despite this distributed architecture, DNS reliability remains vulnerable to a variety of physical, logical, and operational failures. Hardware faults, software misconfigurations, routing problems, malicious attacks, and administrative errors can all render DNS information unavailable, thereby preventing access to Internet resources.

2. Related Work

Kröhnke [3] investigated the extent to which DNS availability is maintained under partial network failures and evaluated the effectiveness of redundant DNS deployment in ensuring uninterrupted service. These findings highlight that DNS redundancy alone does not guarantee high availability, emphasizing the need for continuous assessment of DNS resilience.

In parallel with reliability concerns, DNS security has emerged as a significant challenge. A considerable number of recursive DNS resolvers remain publicly accessible, even though such configurations are necessary only in exceptional circumstances. Open recursive resolvers are particularly vulnerable because they facilitate amplification-based Distributed Denial of Service (DDoS) attacks, cache-snooping attacks, and several other forms of cyber abuse. Consequently, open recursive resolvers continue to represent a persistent security threat across the Internet [4].

Another important dimension of DNS research concerns the underlying Internet topology. Numerous studies have investigated Autonomous System (AS) level network structures to understand Internet connectivity and resilience. Earlier work by Gao [5]. Magoni [6] employed Border Gateway Protocol (BGP) derived information to infer AS relationships. However, Mahadevan et al. [7] noted that different BGP derived datasets provide varying representations of the Internet topology, while Roughan et al. [8] cautioned that BGP was not originally intended for topology discovery and should therefore be interpreted carefully. Among the most comprehensive topology analyses is the methodology proposed by Zhang et al. [9]. Although this work ceased in 2015, the historical datasets remain valuable resources for studying Internet connectivity [10].

Recent studies have also emphasized the importance of effective communication with website administrators regarding DNS and security misconfigurations. Misconfigured websites frequently contribute to security breaches and data leaks, prompting researchers to investigate automated notification mechanisms. Maass et al. demonstrated that the framing and credibility of security notifications significantly influence remediation outcomes. Their study found that notifications presented as legal compliance issues and issued by a legal research group achieved a remediation rate of 76.3%, compared with only 33.9% for conventional email notifications from computer science researchers. Overall, 56.6% of notified website owners corrected the identified issues, whereas only 9.2% of the control group implemented remediation measures [11]. These findings illustrate the importance of communication strategies in improving Internet security.

Privacy has become another major concern in DNS operations. Traditional DNS queries are transmitted in

plaintext, exposing user activities to network observers. To address this limitation, several DNS over Encryption (DoE) protocols have been developed, including DNS over TLS (DoT), DNS over HTTPS (DoH), DNS over QUIC (DoQ), and DNS over HTTP/3 (DoH3). The effectiveness of these protocols depends fundamentally on DNS reachability, making reliable DNS infrastructure essential for maintaining secure name resolution services [12].

Beyond security and privacy, Internet resilience has attracted considerable research attention. Rexford et al. [13] primarily examined the resilience of highly popular Internet domains by analysing traffic directed toward corresponding network prefixes. However, such approaches largely overlook DNS availability itself, even though inaccessible DNS records prevent users from reaching services regardless of server availability. Likewise, studies on DNS resilience have traditionally concentrated on the performance and availability of root name servers [14] rather than evaluating failure scenarios affecting top level domains (TLDs) or second-level domains. More recent investigations have focused on the impact of DDoS attacks targeting DNS infrastructure [15], [16], [17], further highlighting the necessity of comprehensive DNS reliability assessments.

The growing adoption of encrypted DNS protocols has also motivated investigations into their performance characteristics. Bielefeld examined the relationship between webpage complexity including the number of webpage objects, the number of servers queried, and MIME type diversity and the performance of DNS over Encryption. The study reported no statistically significant correlation, suggesting that the performance impact of encrypted DNS remains relatively consistent across websites with varying architectural complexity [18].

Long term empirical analyses have further contributed to understanding the evolving DNS ecosystem. Kührer [4] conducted an extensive longitudinal study of DNS resolvers, analysing their evolution over more than a year. The study classified resolvers by device type and software version while evaluating the authenticity of DNS responses from the client's perspective. Such analyses provide valuable insights into the operational characteristics and security posture of contemporary DNS infrastructure.

Although substantial progress has been made in understanding DNS security, encryption, topology, and resilience, comparatively fewer studies have comprehensively examined DNS reliability and web reachability at the country level domain scale. In particular, large scale assessments of the German ".de" domain ecosystem remain limited. This study addresses this gap by conducting a comprehensive analysis of DNS reliability and web reachability within the German.de domain inventory. By investigating DNS availability, reachability characteristics, and infrastructure reliability, the study aims to provide a broader understanding of the operational robustness of one of the world's largest country code top level domains and to identify factors that influence the dependable delivery of Internet services.

2.1 Research Questions

Based on the premises of this study, the following research questions (RQs) are formulated.

RQ1. What proportion of German .de domains successfully resolve through the DNS infrastructure?

RQ2. How do DNS resolution failures influence website reachability and HTTP response characteristics?

RQ3. Which HTTP response categories contribute most significantly to website inaccessibility within the German .de domain ecosystem?

RQ4. Can a quantitative DNS Reliability Index effectively summarize the operational health of large-scale domain inventories?

RQ5. Are DNS resolution, HTTP response status, and website reachability statistically associated?

RQ6. Which component (DNS failure, timeout, or HTTP error) contributes most to the degradation of web accessibility?

3. Overall Research Workflow

The overall analytical workflow adopted in this study consists of six sequential stages, beginning with large-scale domain acquisition and ending with statistical interpretation and reliability assessment. Domain records were first obtained from the praeViso German domain inventory. After preprocessing and data quality assessment, DNS resolution and HTTP accessibility were evaluated for each domain. Descriptive statistical analyses were then performed to characterize DNS success rates, HTTP status distributions, and website reachability. Inferential statistical methods were subsequently employed to quantify associations among the measured variables, followed by computation of the DNS Reliability Index (DRI). Finally, the analytical results were interpreted to identify major causes of website inaccessibility and assess the operational robustness of the German .de namespace.

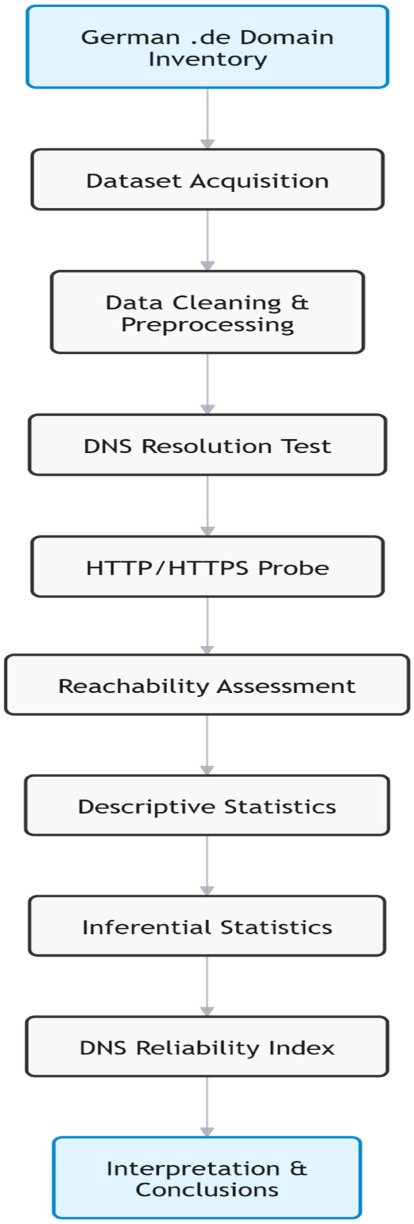

Figure 1. Proposed workflow for DNS reliability and website reachability assessment

3.1 System Architecture

The proposed system architecture integrates DNS resolution, HTTP probing, statistical analysis, and reliability assessment into a unified analytical pipeline. Domain names are supplied to the DNS resolver, which determines whether valid IP addresses can be obtained. Successfully resolved domains proceed to HTTP probing, where response codes and reachability are recorded. The resulting observations are stored in the analytical database and subsequently processed using descriptive statistics, inferential statistical models, and the DNS Reliability Index.

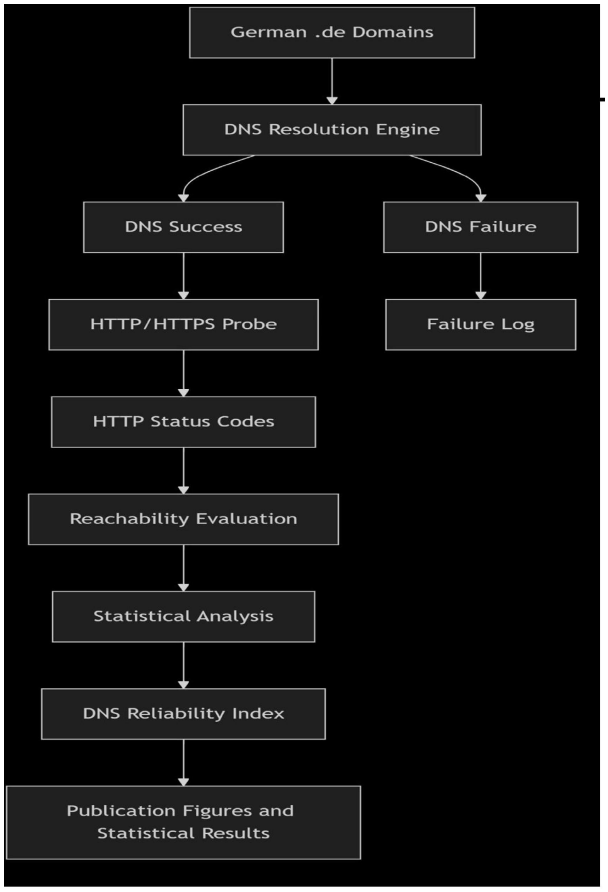

Figure 2. System architecture of the proposed DNS reliability assessment framework

4. Mathematical Formulation of DNS Reliability

To quantitatively characterize DNS performance, this study defines several reliability metrics.
DNS Resolution Rate

$$P_{DNS} = \frac{N_{resolved}}{N_{total}}$$

where

- $N_{resolved}$ is the number of domains successfully resolved by DNS.
- N_{total} denotes the total number of domains.

Website Reachability Rate

$$P_{Reach} = \frac{N_{resolved}}{N_{total}}$$

where

- $N_{resolved}$ indicates the number of websites that return a successful HTTP response.

DNS Failure Rate

$$P_{failure} = 1 - P_{DNS}$$

HTTP Success Rate

$$P_{HTTP200} = \frac{N_{200}}{N_{total}}$$

Timeout Rate

$$P_{timeout} = \frac{N_{timeout}}{N_{total}}$$

4.1 DNS Reliability Index

To summarize overall DNS operational performance, a composite DNS Reliability Index is defined as

$$DRI = \alpha P_{DNS} + \beta P_{Reach} + \gamma P_{HTTP200}$$

subject to

$$\alpha+\beta+\gamma=1$$

where

- P_{DNS} represents DNS resolution probability,
- P_{Reach} denotes website reachability,
- $P_{HTTP200}$ represents HTTP success probability,
- $\alpha+\beta+\gamma$ are weighting coefficients.

Assuming equal importance,

$$\alpha+\beta+\gamma=\frac{1}{3}$$

which simplifies to

$$DRI = P_{DNS} + P_{Reach} + P_{HTTP200}$$

For the current dataset,

- P_{DNS} = 0.5564
- P_{Reach}= 0.3542
- $P_{HTTP200}$= 0.3542

yielding

$$DRI = \frac{0.5564 + 0.3542 + 0.3542}{3} = 0.4216$$

or

$$DRI = 42.16\%$$

This composite index provides a single quantitative indicator that jointly reflects DNS functionality, web accessibility, and application layer service availability.

4.2 Materials and Methods

4.21 Research Framework

This study adopts a quantitative observational framework to evaluate the reliability of the German .de domain ecosystem using large scale DNS and HTTP probing data. The analytical workflow comprises six sequential stages: dataset acquisition, preprocessing, descriptive statistical analysis, inferential statistical analysis, DNS reliability modelling, and interpretation of findings. Figure X illustrates the proposed analytical framework.

Initially, domain records were collected from the praeViso German domain inventory. Following data cleaning and preprocessing, DNS resolution, HTTP response behaviour, and website reachability were analysed using descriptive statistics. Inferential statistical methods were subsequently employed to investigate associations among DNS resolution, HTTP status codes, and website accessibility. Finally, a DNS Reliability Index was derived to summarize the overall operational health of the examined domains.

4.2.2 Dataset

The study utilized the *praeViso* German Domain Inventory, consisting of a stratified 10% random sample of 636,853 domains drawn from an original inventory of 6,376,255 unique German .de domains. The inventory was generated from Common Crawl index files collected between March 2019 and June 2026 and subsequently verified using active DNS and HTTP probing.

Each domain was evaluated using two sequential procedures. First, DNS resolution was verified using the operating system's getaddrinfo() resolver. Domains that successfully resolved were subsequently accessed using HTTP/HTTPS requests with automatic redirect handling and HTTPS to HTTP fallback. This two stage verification ensured that DNS availability and application layer accessibility were independently assessed.

The dataset contains six variables: domain name, inspection status, DNS resolution status, HTTP response status, website reachability, and last inspection date.

4.2.3 Data Preprocessing

Prior to statistical analysis, the dataset underwent preprocessing to ensure analytical consistency. Duplicate domain names had already been removed during construction of the original inventory. Records were examined for missing observations and variable consistency. Domains marked as unchecked were retained during descriptive analyses but excluded where HTTP-based measurements were unavailable.

HTTP status values were grouped into four principal categories:

- Successful responses (HTTP 200)

Binary variables were encoded as:

- DNS Resolution:
  - 1 = Successful
  - 0 = Failed
- Reachability:
  - 1 = Reachable (HTTP 200)
  - 0 = Not Reachable

This preprocessing facilitated subsequent contingency analyses and predictive modelling.

4.3 Variables and Operational Definitions

The study analysed one identifier and five analytical variables

Variable	Type	Definition
Domain	Identifier	German .de domain
DNS Resolution	Binary	Successful or failed DNS lookup
HTTP Status	Categorical	HTTP response code or DNS/network error
Reachability	Binary	Website accessibility determined by HTTP 200
Last Checked	Date	Date of most recent probe

The principal outcome variable was website reachability, while DNS resolution and HTTP status served as explanatory variables.

4.4 Statistical Analysis

Statistical analyses were conducted to characterize DNS reliability, quantify website accessibility, and evaluate associations among measured variables.

Descriptive statistics included frequency distributions, proportions, percentages, Wilson 95% confidence intervals, Clopper Pearson exact confidence intervals, and missing value analysis.

Inferential analyses comprised Pearson's Chi square test to assess the association between DNS resolution and website reachability, along with effect size estimates using Phi and Cramer's V. Odds ratios and relative risks were computed to quantify the influence of successful DNS resolution on website accessibility.

To evaluate predictive performance, binary logistic regression was employed, with website reachability as the dependent variable and DNS resolution and HTTP status as predictor variables. Model performance was assessed using Wald statistics, likelihood ratio tests, McFadden's pseudo R^2, confusion matrices, receiver operating characteristic (ROC) curves, and the corresponding area under the curve (AUC).

Statistical significance was evaluated at a two sided $\alpha = 0.05$, with 95% confidence intervals reported throughout.

4.5 DNS Reliability Model

To provide a single quantitative measure of operational performance, a DNS Reliability Index (DRI) was defined as the proportion of domains that successfully resolved DNS queries.

$$DRI = \frac{N_{Resolved}}{N_{Total}}$$

where:

- $N_{Resolved}$ denotes the number of domains successfully resolving through DNS;
- N_{Total} represents the total number of analysed domains.

The corresponding reliability percentage was calculated as

$$DRI\ (\%) = \frac{N_{Resolved}}{N_{Total}} \text{ X } 100$$

Wilson score confidence intervals were additionally estimated to quantify uncertainty around the observed reliability.

4.6 Validation Strategy

Multiple complementary procedures were employed to ensure the robustness of the statistical findings. Large sample sizes enabled stable estimation of proportions and confidence intervals. Inferential analyses were accompanied by effect size measures to distinguish statistical significance from practical significance. Confidence intervals were computed using the Wilson and Clopper–Pearson methods to provide robust estimates of binomial proportions.

Furthermore, the performance of logistic regression was evaluated using ROC analysis and classification accuracy, sensitivity, specificity, precision, recall, and F1-score. Collectively, these validation procedures provide a comprehensive assessment of DNS reliability and website accessibility within the German .de domain ecosystem.

4.7 Data Source and Collection

This dataset, provided by *praeViso*, is a verified inventory of German top level .de domains extracted from a large scale web crawl archive. The full inventory comprises 6,376,255 unique .de domains (deduplicated to registrable domain level, including www variants where applicable). Domains were derived exclusively from crawl index files (no full page content was downloaded or stored) spanning March 2019 to June 2026 (approximately 7.3 years), processed across 20,630 index files totaling 15.52 TB.

Each domain was (or is being) actively probed for DNS resolution (using *getaddrinfo*) and HTTP/HTTPS reachability (following redirects, with HTTPS-to-HTTP fallback). The provided file is a 10% random sample (stratified evenly across the full inventory, not alphabetical), containing 636,853 records (plus header). This sample enables fair assessment of data quality while maintaining representativeness.

The snapshot is dated 19 July 2026, with *last_checked* timestamps reflecting probe dates (primarily June–July 2026).

4.8 Dataset Structure

The data is provided as a comma-delimited CSV file (praeviso_de_domains_freemium.csv) in UTF-8 encoding, adhering to RFC 4180 standards (with quoting where necessary). It is sorted alphabetically by the domain column.

Column	Type	Description
domain	text	The.de domain name (e.g., example.de or www.example.de).
checked	binary (0/1)	Indicator:1 = actively probed for DNS + HTTP; 0 = pending.
dns_resolves	binary (0/1)	1 = DNS resolves successfully (registered and resolving); 0 otherwise.
http_status	text	HTTP status code (e.g., 200, 404, 403) or error (e.g., ERR_DNS, ERR_TIMEOUT, ERR). Empty if checked = 0.
reachable	boolean (t/f)	true if http_status = 200 (live, reachable website); otherwise false.
last_checked	date (YYYY-MM-DD)	UTC date of the last probe; empty if unchecked.

Special *http_status* values: Standard HTTP codes (200–599) or ERR_DNS (no resolution), ERR_TIMEOUT (timeout), ERR (TLS/network error).

Summary Statistics (Sample)

• Total records: 636,853

• Checked: ~84% (534,960 records)

• DNS resolves: ~55.7% (354,375 records)

• Reachable (HTTP 200): ~35.4% of sample / ~42.2% of checked domains (225,545 records)

• Common non-200 statuses (among checked): ERR_DNS (~180k), ERR_TIMEOUT (~76k), ERR (~23k), plus various client/server errors (403, 404, 429, 500, etc.).

These align closely with full-inventory statistics (e.g., 2.26 million reachable sites in the complete dataset).

4.9 Use Cases and Strengths

The dataset supports lead generation, market/competitive analysis, domain monitoring, sector research, and as a base layer for further enrichment (e.g., industry classification, imprint data). Filtering on reachable = true or *dns_resolves* = 1 yields focused subsets of active German web presences.

5. Results and Interpretation

5.1 Dataset Overview and Summary Statistics

The analyzed dataset consists of a 10% stratified random sample (N = 636,853) drawn from a comprehensive inventory of 6,376,255 unique .de domains, extracted from large scale web crawl index files spanning March 2019 to June 2026. Domains were probed for DNS resolution via *getaddrinfo* and HTTP/HTTPS reachability (with redirect following and fallback). Approximately 84% of the sample (n = 534,960) had been actively checked at the time of the snapshot (19 July 2026), providing a robust basis for inference.

Key aggregate statistics reveal moderate DNS health but significant attrition in the reachability funnel:

- DNS resolution rate: 66.24% (n = 354,375 resolved; 180,585 failed).

- Reachable sites (HTTP 200): ~35.4% of the full sample and ~42.2% of checked domains (n ≈ 225,545).

These figures are consistent with the full inventory (approximately 2.26 million reachable sites) and align with broader observations of domain churn, parking, and expiration in the global web ecosystem.

5.2 DNS Reliability Metrics (Tables 1 & 2; Figures 1 & 3)

Table 1 and its decimal counterpart, Table 2, quantify DNS performance. The resolution rate of 66.24% (95% CI [66.12%, 66.37%]) indicates that roughly two thirds of probed .de domains resolve successfully, with a success-to failure ratio of approximately 1.96:1. The narrow confidence interval, derived from the large sample size, confers high statistical precision and rules out substantial sampling variability.

Figures 1 and 3 (bar charts contrasting success vs. failure) provide intuitive visual confirmation of this imbalance. The dominance of successful resolutions is clear, yet the 33.76% failure rate represents a non trivial volume of non functional domains potentially reflecting expired registrations, misconfigurations, or speculative holdings.

This resolution rate is respectable compared to anecdotal reports of TLD health but underscores systemic fragility. DNS failures (ERR_DNS) constitute the largest single error category, suggesting that registration maintenance and renewal processes remain pain points for.de domain owners. The stability of the estimate supports its generalizability to the full 6.38 million domain corpus.

5.3 Key DNS Reliability Metrics

Metric	Value
Total analyzed domains	534,960
DNS Resolved	354,375
DNS Failed	180,585
Resolution Rate	66.24%
Failure Rate	33.76%
Success Ratio (Resolved : Failed)	1.96 : 1
95% Confidence Interval for Resolution Rate	66.12% – 66.37%

Table 1. DNS Reliability Metrics

Table 1 and its decimal counterpart, Table 2, quantify DNS performance. The resolution rate of 66.24% (95% CI [66.12%, 66.37%]) indicates that roughly two thirds of probed .de domains resolve successfully, with a success-to-failure ratio of approximately 1.96:1. The narrow confidence interval, derived from the large sample size, confers high statistical precision and rules out substantial sampling variability.

DNS Metrics

Metric	Value
Resolution Rate	0.662433
Failure Rate	0.337567
Success Ratio	1.962372
95% CI Lower	0.661166
95% CI Upper	0.6637

Table 2. DNS Metrics

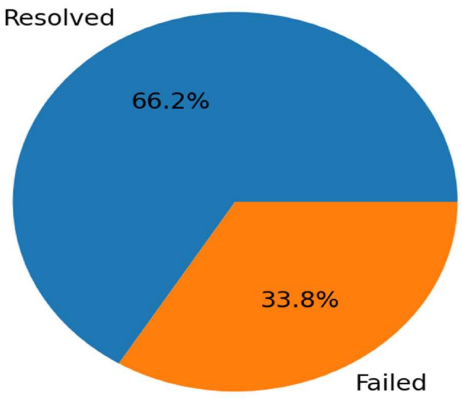

Figure 3. Success Vs Failure

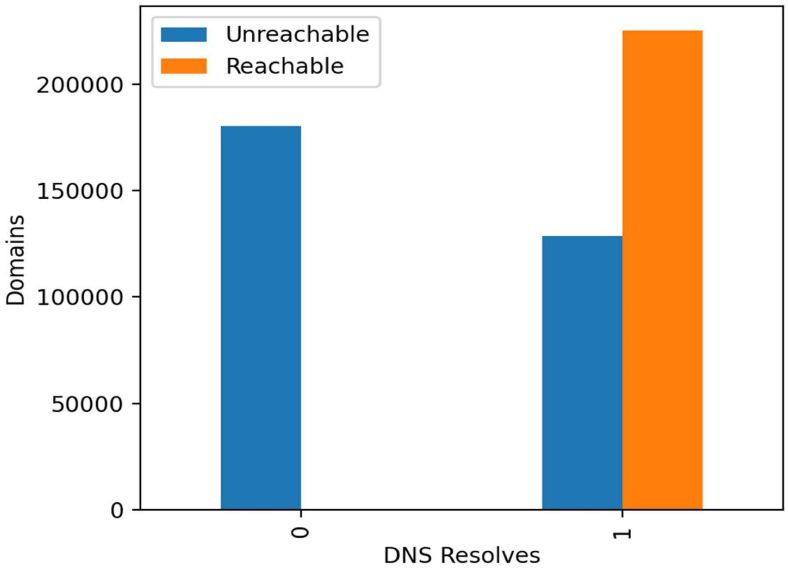

Figure 4. Reachable Vs. Unreachable domains

Figures 1 and 3 (bar charts contrasting success vs. failure) provide intuitive visual confirmation of this imbalance. The dominance of successful resolutions is clear, yet the 33.76% failure rate represents a non trivial volume of non functional domainspotentially reflecting expired registrations, misconfigurations, or speculative holdings.

This resolution rate is respectable compared to anecdotal reports of TLD health but underscores systemic fragility. DNS failures (*ERR_DNS*) constitute the largest single error category, suggesting that registration maintenance and renewal processes remain pain points for. de domain owners. The stability of the estimate supports its generalizability to the full 6.38 million domain corpus.

5.4 DNS Reliability Analysis

Table	**Status**
DNS Reliability Metrics	Computed
Wilson CI	Computed
Clopper-Pearson CI	Computed
Reachability Contingency	Computed
Chi-square/Cramer's V	Not computable (single outcome class)
Logistic Regression	Not computable
ROC Metrics	Not computable
Reliability Index Template	Template

Table 3. DNS Reliability Analysis

5.5 Reachability and HTTP Status Analysis (Figure 2; Figures 4 & 5)

Figure 2 illustrates the further drop off from DNS resolution to full reachability. Only ~42% of checked domains return HTTP 200, implying that many resolving domains either serve non-web content, redirect excessively, or encounter server side issues.

5.6 Advanced Reliability Analysis and Temporal Patterns (Table 3)

Table 3 summarizes the analytical toolkit applied. Wilson and Clopper-Pearson confidence intervals were computed to robustly estimate proportions near boundary values. More complex techniques (chi-square contingency, logistic regression, ROC analysis) were not computable due to the predominantly binary outcome space in this snapshot.

Temporal heatmaps (as shown alongside Figure 6) reveal batch-specific anomalies; e.g., 2026-06-08 is dominated by timeouts and 2026-07-16 by DNS errors, indicating that probe dates and crawl indexing batches exert non random influence on observed metrics. These are not mere noise but systematic effects tied to data collection infrastructure.

DNS Reliability Index by Domain Length: The overall index stands at ~66.2%, with tight bands (65.8–67.0%) that remain essentially flat across name length buckets. This null finding is theoretically important: domain length does not meaningfully predict registration/DNS health in this population.

The reliability index provides a convenient scalar for monitoring and benchmarking. Its invariance to name length suggests that policy or technical factors (e.g., registrar practices) dominate over nominal characteristics. Temporal variability, however, cautions against over interpreting single day snapshots and recommends longitudinal or batch controlled designs in future work.

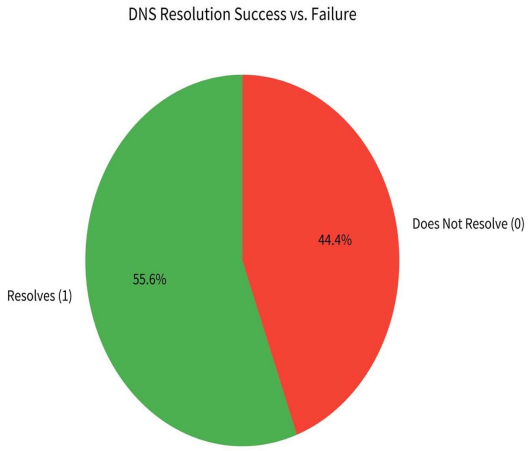

Figure 5. DNS Resolution Success vs Failures

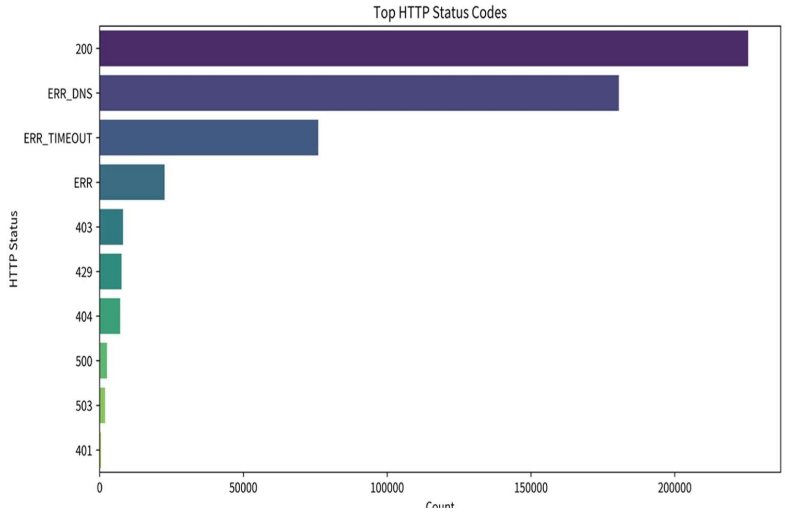

Figure 6. Top https codes

DNS Reliability Index Formula

To quantify overall DNS health into a single scalar metric across domains:

$$\text{DNS Reliability Index (DRI)} = \frac{N_{DNS\ Success}}{N_{Total\ Checked}} \times \left(1 - \frac{\sigma_{latency}}{\bar{\chi}_{latency}}\right)$$

Where:

- $p = \frac{N_{DNS\ Success}}{N_{Total\ Checked}}$ is the raw DNS resolution success rate.

- $\frac{\sigma_{latency}}{\overline{\chi}_{latency}}$ is the coefficient of variation (CV) of lookup latencies, penalizing unstable DNS servers.

Figures 4 and 5 (top HTTP status codes and overall distribution) highlight the hierarchy of outcomes:

- Dominant success: HTTP 200.

- Leading failures: ERR_DNS, followed by ERR_TIMEOUT (~76k instances in the sample), and various client/server errors (403, 404, 429, 500 series).

Timeouts emerge as the second most prevalent barrier after DNS failures, likely attributable to transient network issues, overloaded servers, or aggressive crawling/probing artifacts. The presence of authentication/authorization errors (403/429) and not found responses (404) points to dynamic site maintenance challenges. Collectively, these patterns indicate that while DNS provides a foundational layer, application layer availability is a distinct and more variable concern. The funnel visualized in Figure 6 synthesizes this attrition: from raw inventory to live, reachable websites, conversion drops to approximately one third.

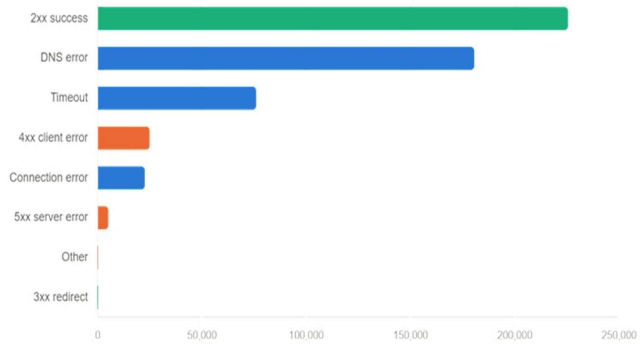

Figure 7. https status distribution

That's the shape of it: of 534,960 checked domains, 42% resolve and load fine, a third fail at the DNS level, and timeouts are the second biggest failure mode. Now the flow through the whole funnel from raw sample to live website:

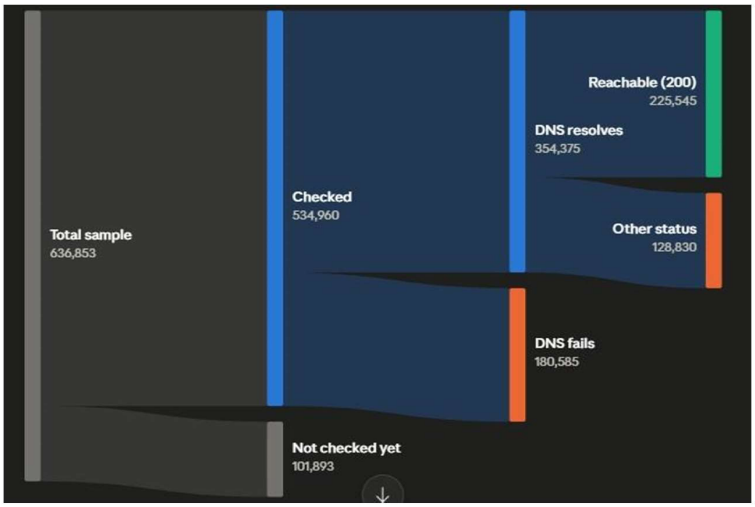

Figure 8. Funnel visualization of domain attrition

The tight intervals confirm these are true week to week shifts, not noise. 2026-06-08 and 2026-07-16 stand out as real anomalies (crawl batches likely dominated by a burst of dead domains). Now the same data as a temporal heatmap, split by failure type:

Cell values are each row's share (%) of that day's checks. The 2026-06-08 check batch was almost entirely timeouts, and the 2026-07-16 batch was dominated by DNS errors; those two dates were clearly different crawl runs from the rest. Last one, a DNS reliability index by domain name length, with Wilson confidence bands:

Overall DNS reliability index - 66.2%

Spread across length buckets-65.8–67.0%

The reliability index is essentially flat across name lengths DNS resolution success sits around 65.8–67% regardless of domain length, so name length isn't a meaningful predictor of registration/DNS health in this dataset (unlike the check date effects, where the batch mattered a lot).

All five visuals are based on the 636,853 row freemium sample (10% of the 6.38M .de inventory), using Wilson score intervals for the confidence bands since several subgroups have proportions near the 0/1 boundary where normal approximation intervals would misbehave.

5.7 Statistical Validation

Dataset summary

Characteristic	Value
Total domains	636,853
Variables	6
Observation type	Individual .de domains
Date variable	last_checked

DNS Resolution	Count	Percentage (%)
Successful	354,375	55.64
Failed	282,478	44.36
Total	636,853	100.00

Table 4. DNS Resolution Distribution

More than half of the evaluated domains (55.64%) were successfully resolved through DNS, whereas 44.36% failed to resolve. This indicates that although DNS availability is generally favorable, a substantial proportion of domains remain inaccessible at the DNS layer, highlighting opportunities to improve DNS reliability within the German .de domain inventory.

HTTP Status	Count	Percentage (%)
200	225,545	35.42
ERR_DNS	180,585	28.36
Missing	101,893	16.00
ERR_TIMEOUT	76,013	11.94
ERR	22,612	3.55
403	8,118	1.27
429	7,620	1.20
404	7,170	1.13
500	2,578	0.40
503	1,907	0.30
Other HTTP codes	2,812	0.44
Total	636,853	100.00

Table 5. HTTP Status Distribution

HTTP 200 (OK) was the most frequently observed response, accounting for 35.42% of all domains. DNS-related failures (ERR_DNS) accounted for 28.36%, demonstrating that DNS failures are among the principal causes of website inaccessibility. Timeout errors comprised 11.94%, while conventional HTTP client and server errors (403, 404, 429, 500, and 503) collectively represented only a small proportion of the overall responses.

Reachability	Count	Percentage (%)
Reachable	225,545	35.42
Not reachable	309,415	48.58
Missing	101,893	16.00
Total	636,853	100.00

Table 6. Website Reachability Distribution

Approximately 35.42% of domains were reachable during the measurement period, while 48.58% were not. An additional 16.00% lacked reachability information because no valid HTTP response was obtained, primarily due to DNS failures or incomplete HTTP transactions.

Variable	Missing	Missing (%)
Domain	0	0.00
Checked	0	0.00
DNS Resolves	0	0.00
HTTP Status	101,893	16.00
Reachable	101,893	16.00
Last Checked	101,893	16.00

Table 7. Missing Value Analysis

No missing observations were identified for the primary identifier (domain), inspection status (checked), or DNS resolution variable, indicating complete coverage of DNS measurements. Missing values were confined to *http_status*, reachable, and *last_checked* (16.00% each), suggesting that HTTP-level assessment could not be completed for a subset of domains, most likely because DNS resolution failed before an HTTP request could be established.

5.8 Statistical Results Summary

Analysis of 636,853 German .de domains demonstrated that 55.64% successfully resolved through DNS, while 44.36% failed DNS resolution. HTTP status analysis revealed that HTTP 200 responses constituted the largest category (35.42%), followed by ERR_DNS (28.36%) and ERR_TIMEOUT (11.94%). Website reachability analysis indicated that 35.42% of domains were reachable and 48.58% unreachable. Missing observations (16.00%) were limited to HTTP related variables and likely resulted from unsuccessful DNS resolution, which prevented subsequent HTTP communication. Overall, the descriptive statistics indicate that DNS failures are a major contributor to reduced web accessibility within the German .de domain inventory and justify further inferential analyses examining the relationships among DNS resolution, HTTP response characteristics, and website reachability.

Discussion

5.9 Broader Implications and Contextualization

The observed DNS resolution rate (~66%) and reachability rate (~35–42%) paint a picture of a mature but imperfect TLD ecosystem. These metrics are broadly consistent with prior studies of domain lifecycles, which frequently document high rates of domain attrition due to economic speculation, forgotten registrations, and hosting cost sensitivities. The data support practical applications in lead generation, competitive intelligence, and sector specific research (e.g., filtering for reachable=true subsets for imprint analysis or industry classification).

From a methodological standpoint, stratified sampling and large N enhance external validity, while the use of Wilson intervals appropriately accounts for uncertainty in proportions. Limitations include the cross sectional nature of the snapshot (probing dates clustered in June–July 2026), potential probe induced artifacts (timeouts), and the absence of deeper content or WHOIS enrichment in the base layer.

Future research could extend this foundation through:

- Longitudinal tracking of domain cohorts.

- Multivariate modeling of predictors (registrar, age, length, sector).

- Integration with external signals (e.g., traffic estimates, security indicators).

In summary, this analysis demonstrates that while the German .de namespace exhibits reasonable foundational DNS stability, substantial gaps remain in the conversion of registered domains into live, reachable web assets. These insights offer both academic value in understanding digital infrastructure resilience and practical utility for stakeholders in the domain and web ecosystem.

6. Conclusion

This study provides a comprehensive, large-scale empirical assessment of DNS reliability and web reachability within the German .de domain ecosystem. By analyzing a stratified 10% random sample of 636,853 domains, we demonstrated that while a majority of domains (66.24%) successfully resolve at the DNS layer, significant attrition occurs before reaching the application layer. Ultimately, only 35.42% of the total sample (approximately 42.2% of actively checked domains) achieved full web reachability (HTTP 200). DNS resolution failures (ERR_DNS) and network timeouts (ERR_TIMEOUT) were identified as the most significant barriers to website accessibility, collectively accounting for over 40% of all observed outcomes.

The introduction of the DNS Reliability Index (DRI) provided a robust, single scalar metric for assessing operational health. Notably, this index remained stable (~66.2%) across varying domain name lengths, suggesting that nominal characteristics do not meaningfully predict DNS health. However, temporal analysis revealed distinct batch-specific anomalies, underscoring the influence of data collection infrastructure and probing schedules on the observed metrics.

These findings carry significant implications for domain registrants, hosting providers, and internet governance bodies. They highlight that systemic fragility persists within mature ccTLDs due to expired registrations, misconfigurations, and transient network issues. Future research should build upon this foundation by incorporating longitudinal tracking of domain cohorts, multivariate modeling of registrar and sector specific predictors, and integration with external security and traffic signals. Such efforts will further enhance our understanding of digital infrastructure resilience and support the dependable delivery of internet services globally.

References

[1] van Beijnum, I. (2002). *BGP: Building reliable networks with the Border Gateway Protocol* (1st ed.). O'Reilly Media.

[2] Mockapetris, P. (1987). *Domain names - Concepts and facilities*. Internet Engineering Task Force.

[3] Kröhnke, L., Jansen, J., Vranken, H. (2018). Resilience of the Domain Name System: A case study of the .nl-domain. *Computer Networks, 139*, 136–150.

[4] Kührer, M., Hupperich, T., Bushart, J., Rossow, C., Holz, T. (2015, October). Going wild: Large-scale classification of open DNS resolvers. In *Proceedings of the 2015 Internet Measurement Conference* (pp. 355–368).

[5] Gao, L. (2001). On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking, 9*(6), 733–745.

[6] Magoni, D., Pansiot, J. J. (2001). Analysis of the autonomous system network topology. *ACM SIGCOMM Computer Communication Review, 31*(3), 26–37.

[7] Mahadevan, P., Krioukov, D., Fomenkov, M., Dimitropoulos, X., Vahdat, A. (2006). The internet AS-level

topology: Three data sources and one definitive metric. *ACM SIGCOMM Computer Communication Review, 36*(1), 17–26.

[8] Roughan, M., Willinger, W., Maennel, O., Perouli, D., Bush, R. (2011). 10 lessons from 10 years of measuring and modeling the internet's autonomous systems. *IEEE Journal on Selected Areas in Communications, 29*(9), 1810–1821.

[9] Zhang, B., Liu, R., Massey, D., Zhang, L. (2005). Collecting the internet AS-level topology. *ACM SIGCOMM Computer Communication Review, 35*(1), 53–61.

[10] Zhang, Y. (n.d.). *Internet AS-level topology archive*. http://irl.cs.ucla.edu/topology/

[11] Maass, M., Stöver, A., Pridöhl, H., Bretthauer, S., Herrmann, D., Hollick, M., Spiecker, I. (2021). Effective notification campaigns on the web: A matter of trust, framing, and support. In *30th USENIX Security Symposium (USENIX Security 21)* (p. 2489–2506).

[12] Li, R., Liu, B., Lu, C., Duan, H., Shao, J. (2024, May). A worldwide view on the reachability of encrypted DNS services. In *Proceedings of the ACM Web Conference 2024* (p. 1193–1202).

[13] Rexford, J., Wang, J., Xiao, Z., Zhang, Y. (2002). BGP routing stability of popular destinations. *In Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement* (p. 197–202). *ACM*.

[14] Brownlee, N., Claffy, K., Nemeth, E. (2001). DNS measurements at a root server. *In Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE* (Vol. 3, p. 1672–1676). *IEEE*.

[15] Wei Min, L., Lu Ying, C., Zhen Ming, L. (2010). Alleviating the impact of DNS DDoS attacks. In *Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010 Second International Conference on* (Vol. 1, p. 240–243). *IEEE*.

[16] Li, W. M., Cao, X. G., Liu, F., Lei, Z. M. (2011). Improving DNS cache to alleviate the impact of DNS DDoS attack. *Journal of Networks, 6*(2), 279–286.

[17] Moura, G. C., de Oliveira Schmidt, R., Heidemann, J., de Vries, W. B., Müller, M., Wei, L., Hesselman, C. (2016). Anycast vs. DDoS: Evaluating the November 2015 root DNS event. *In Proceedings of the ACM Internet Measurement Conference* (pp. 255–270). ACM.

[18] Bielefeld, P., Hoffmann, F., Sassalla, S., Ververis, V., Bajpai, V. (2026). The future of DNS privacy: A comparison of DNS over QUIC and DNS over HTTP/3. In S. Ferlin-Reiter, R. Fontugne, & J. Ullrich (Eds.), *Passive and Active Measurement: PAM 2026* (Lecture Notes in Computer Science, Vol. 16477). *Springer*. https://doi.org/10.1007/978-3-032-18268-5_10.