1. Primary Cohort Germany (n = 10,648 reachable homepages)

Inspection of homepages from public directories (Measurement period: Sept. - Oct. 2026)

Examined Characteristic	Denominator	Share (%)	Absolute Count
Postal address in visible text / legal notice	n = 10,648 (Total cohort)	**94.1 %**	ca. 10,020
No Schema.org business markup in JSON-LD	n = 10,648 (Total cohort)	**73.0 %**	7,778
Schema.org business markup present	n = 10,648 (Total cohort)	**27.0 %**	2,870
- thereof nested PostalAddress	n = 2,870 (Subgroup with markup)	**21.3 %**	611
- PostalAddress relative to total cohort	n = 10,648 (Total cohort)	**6.5 %**	695
Valid /llms.txt file present (soft-404 filtered)	n = 10,648 (Total cohort)	**4.7 %**	497
/llms.txt before soft-404 filtering (raw)	n = 10,648 (Total cohort)	**9.7 %**	1,030
AI training bots blocked (GPTBot etc.)	n = 10,648 (Total cohort)	**4.3 %**	453
AI search crawlers blocked (Perplexity etc.)	n = 10,648 (Total cohort)	**2.2 %**	239
- Open for AI search crawlers	n = 10,648 (Total cohort)	**97.8 %**	10,409
Global wildcard disallow (Disallow: /)	n = 10,648 (Total cohort)	**0.6 %**	60

2. International Country Comparison (Germany, UK/EN, USA, Total)

Comparison of reachable homepages across different top-level domains and country codes

Market / Segment	Reachable (n)	Avg. Score	Schema Markup	PostalAddress	Search Bots Open
Germany (DE)	n = 10,648 (63.8 %)	**65.6 / 100**	27.0 % (2,870)	6.5 % (695)	97.8 % (10,409)
UK / English-speaking (EN)	n = 839 (59.3 %)	**74.4 / 100**	54.1 % (454)	28.5 % (239)	98.5 % (826)
United States (US)	n = 2,114 (29.2 %)	**73.2 / 100**	53.8 % (1,138)	29.7 % (627)	98.6 % (2,085)
Other Markets / International	n = 109 (51.9 %)	**69.3 / 100**	31.2 % (34)	21.1 % (23)	96.3 % (105)
Total (All Markets)	n = 13,710 (53.6 %)	**67.4 / 100**	32.8 % (4,496)	11.6 % (1,584)	97.9 % (13,425)

Methodological Framework & Standards Applied (W3C & RFC):

- Data Source & Sample Integrity: Homepages of reachable corporate websites from public directories (n = 13,710 globally, including n = 10,648 in Germany; measurement period: September - October 2026).
- Strict Pre-Filtering: Domains with DNS, TLS, or HTTP connection failures were excluded prior to analysis. All percentage figures refer strictly to verified, functioning websites.
- Standards Applied: Schema.org Vocabulary v26.0 (Organization, LocalBusiness, PostalAddress), W3C JSON-LD 1.1 Recommendation, RFC 9309 (Robots Exclusion Protocol), llms.txt standard.
- /llms.txt Methodology: Rigorous distinction between superficial HTTP 200 availability and actual Markdown content after filtering soft-404 error responses (valid Markdown file with H1 title).

3. Regional Economic Clusters & Metropolitan Areas in Detail

DATA SHEET FOR REGIONAL & BUSINESS EDITORIAL TEAMS (MÜNSTERLAND, OSNABRÜCK, METROS)

A) Regional Economic Clusters (Germany)

Economic Region	Sample (n)	Avg. Score	PostalAddress (JSON-LD)	Gap Without Address
Münsterland (Rheine, MS, ST)	n = 228 (94.6 %)	**68.6 / 100**	12.7 % (29)	**87.3 % (199)**
Osnabrück Region (City & District)	n = 183 (98.9 %)	**60.4 / 100**	4.4 % (8)	**95.6 % (175)**
Euregio (Bentheim, Emsland)*	n = 46 (100.0 %)	**68.7 / 100**	6.5 % (3)	**93.5 % (43)**

** Euregio Note: Cross-border pilot sample (n = 46).*

B) National & International Metropolitan Areas

Metropolitan Area	Sample (n)	Avg. Score	PostalAddress (JSON-LD)	/llms.txt Present
Berlin Metropolitan Area (DE)	n = 256 (100.0 %)	**71.1 / 100**	21.5 % (55)	13.7 % (35)
Munich Metropolitan Area (DE)	n = 252 (98.8 %)	**70.4 / 100**	18.3 % (46)	13.1 % (33)
Hamburg Metropolitan Area (DE)	n = 280 (98.9 %)	**69.3 / 100**	10.4 % (29)	7.9 % (22)
Rhine-Ruhr Metro (Cologne, Düsseldorf)	n = 958 (99.6 %)	**69.1 / 100**	14.0 % (134)	9.2 % (88)
Frankfurt Rhine-Main (DE)	n = 310 (98.1 %)	**67.9 / 100**	13.5 % (42)	8.1 % (25)
Greater New York (US Metro)	n = 59 (98.3 %)	**82.3 / 100**	69.5 % (41)	30.5 % (18)
San Francisco Bay Area (US)	n = 47 (94.0 %)	**82.2 / 100**	85.1 % (40)	48.9 % (23)
Greater London (UK Metro)	n = 43 (87.8 %)	**88.7 / 100**	97.7 % (42)	39.5 % (17)

*** Note: Metro values report verified valid /llms.txt files after soft-404 filtering.*

Key Regional Findings & Actionable Recommendations:

- Münsterland & Rheine: 94.6 % of examined domains were technically reachable, but 87.3 % omit structured Schema.org location data in their source code. Machine-readable address data is an underestimated lever, particularly for regional crafts and medium-sized enterprises.
- Osnabrück Region: With an address gap of 95.6 %, many businesses miss the opportunity to be unambiguously identified in location-based AI search queries (e.g. 'mechanical engineering Osnabrück').
- Metropolitan Comparison: While structured address entities are standard in US and UK tech hubs (San Francisco 85.1 %, London 97.7 %), Berlin lags behind at 21.5 % (55 of 256; missing on 78.5 % / 201; focused cohorts, non-representative).

Concrete Steps for Webmasters, SMEs, and Regional Businesses:
1. Embed structured address data: Add schema.org/PostalAddress inside LocalBusiness or Organization JSON-LD.
2. Differentiate AI crawler access: Permit real-time retrieval bots (such as OAI-SearchBot) in robots.txt.
3. Provide valid /llms.txt overview: Implement a concise Markdown file without HTML error wrappers for AI agents.

Press Contact, Data Material & Terms of Use:

MIRDAT - Michael Röttger | Friedrich-Ebert-Ring 154 | 48429 Rheine, Germany
Email: info@mirdat.de | Central Source Page & Factsheet Download: https://www.mirdat.de/en/press
Citation: "Source: MIRDAT Factsheet Website Study 2026 (mirdat.de/en/press)" | Graphics & tables free under CC BY 4.0