

What can a machine verify on Spain's public-sector websites?

EasyxLab · EasyByte Hub S. Coop. Mad. · Study S8

Working draft Published 2026-10-03

Abstract

Several obligations and good practices for public websites leave a trace that a program can check without human judgement: a `security.txt` file (RFC 9116), an accessibility statement that is linked from the site and "actualizada periódicamente, como mínimo una vez al año" (Spain's Royal Decree 1112/2018, art. 15.1), HTTPS with HSTS, and a stated policy towards AI crawlers in `robots.txt`. We checked these on the websites of 6,648 of Spain's 8,132 municipalities (81.8% of municipalities, 98.9% of the population), plus the 19 regional governments, 52 provincial and island councils, 50 public universities and 22 ministries. Of 6,730 entities whose home page we could measure, 5,245 (77.9%) returned their own home page; the rest failed, refused our client, were closed to us by their `robots.txt`, or led to something that is not the entity's site (179 URLs: hijacked domains, hosting panels, parked domains). Of the 5,245, 5,130 (97.8%) served the home page over HTTPS when asked for it, and 1,688 of those 5,130 (32.9%) sent HSTS. Of 5,570 entities whose server answered a request for `/.well-known/security.txt`, 12 (0.2%) serve a file, 8 have its two required fields and have not expired, and 4 meet every requirement of RFC 9116 we can test, while 329 (5.9%) answer that path with HTTP 200 and something else. 2,492 of the 5,245 home pages (47.5%) carry a link to the accessibility statement that our detector can see; of 1,247 statements read with an audited date extractor, 908 (72.8%) show a preparation or review date and 160 (12.8%; 95% CI 11.1–14.8) show one from the last 365 days. Spain's official monitoring already checks, by experts on a sample of about 63 websites a year, whether a statement is provided and what it says (§2); it does not report whether home pages link to it or how old its date is. The six most frequent dates each occur in a single province and together account for 479 of the 908 dated statements (52.8%). 326 of 6,012 entities (5.4%) block at least one of GPTBot, ClaudeBot, Google-Extended or CCBot, concentrated in a few regions; a spec-like `llms.txt` appears on 7 of 1,020 sites in a random subsample (0.7%). A first version of our scanner fetched the home page of 123 entities, and `security.txt`, `llms.txt` or the statement of another 722, although their `robots.txt` did not allow it; those records were deleted, and the 123 are reported as not measurable (§5.3). We publish the per-entity table, the cleaned scan records, the scanner and the method.

1. Introduction

Compliance with web obligations in Spain's public sector is usually assessed by sampling and by people: accessibility monitoring works on samples of sites, and audits read statements and policies. Some of the requirements, though, are mechanical. Either a file exists at a fixed path or it does not; either a header is sent or it is not; either a page shows a date less than a year old or it does not. Those facts can be collected for every public body at almost no cost.

This study asks how far a census of such facts can go for the Spanish public sector, what it finds, and where it stops. We restrict ourselves to checks that need no judgement: we do not assess whether a site is accessible, whether a statement is truthful, or whether a security contact answers. We also measure the measurement: how many public bodies have a findable website at all, how many published URLs no longer lead to the body's site, and how reliable the one non-trivial step (reading a date from a page) is.

2. Prior work

How we searched. On 2026-10-02, between 22:20 and 22:51 UTC, the agent queried:

- the **arXiv API** (`export.arxiv.org/api/query` , 25 results per query, by relevance):
`all:"security.txt" ; abs:HSTS AND abs:government ; abs:"government websites" AND abs:HTTPS ; abs:"accessibility statement" ; abs:"accessibility statements" ; abs:robots.txt AND abs:AI AND abs:crawlers ; all:"llms.txt" ; abs:municipal AND abs:websites AND abs:Spain ; abs:"public sector" AND abs:websites AND abs:accessibility ; abs:"vulnerability disclosure" AND abs:security.txt ; ti:"Accept the Risk and Continue" ; abs:"Consent in Crisis" ; abs:robots.txt AND abs:"AI" ; abs:"government" AND abs:"HTTPS adoption" ; abs:"web accessibility" AND abs:"municipalities" ;`
- the **Crossref REST API** (`api.crossref.org/works` , `query.bibliographic` , 10–15 results per query):
"security.txt vulnerability disclosure adoption"; "security.txt prevalence conformity"; "security.txt revisited prevalence"; "security.txt files websites analysis"; "HTTPS adoption government websites"; "Accept the Risk and Continue government https adoption"; "HSTS deployment measurement"; "Strict-Transport-Security government"; "accessibility statements public sector websites"; "accessibility statement Web Accessibility Directive compliance"; "web accessibility directive monitoring"; "web accessibility Spanish municipalities"; "accesibilidad web ayuntamientos"; "declaración de accesibilidad sitios web"; "robots.txt AI crawlers measurement";
- the **Semantic Scholar Graph API** (`paper/search` , 15 results per query): answered "security.txt adoption", "government websites HTTPS adoption measurement", "accessibility statements Web Accessibility Directive public sector websites", "accesibilidad web ayuntamientos españoles" and "Accept the Risk and Continue government https"; refused with HTTP 429 (rate limit), after four tries each, "security.txt vulnerability disclosure websites measurement", "HSTS deployment government websites", "Spanish municipalities websites accessibility", "robots.txt AI crawlers blocking measurement" and "llms.txt adoption";
- **GitHub**, with `gh search repos` (8 results per query): "security.txt government", "security.txt scanner", "securitytxt", "security.txt in:name", "pshtt", "HSTS government websites", "government https scan", "pulse https government", "domain-scan hsts", "gov websites security headers", "site-scanning", "declaración de accesibilidad", "accesibilidad ayuntamientos", "accesibilidad web in:readme", "observatorio accesibilidad web", "OAW accesibilidad", "rastreador accesibilidad", "ayuntamientos web scraping", "spain municipalities websites", "robots.txt AI crawlers dataset", "ai robots.txt", "llms.txt checker";
- the **official monitoring of RD 1112/2018**: the pages of the Observatorio de Accesibilidad Web (OAW) on the Portal de Administración Electrónica, which publish the annual monitoring results (2020–2021 to 2025) and Spain's reports to the European Commission under Directive (EU) 2016/2102 (2020–2021 and 2022–2024); and the Commission's pages listing the national monitoring reports.

General web search was not available. Grey literature beyond these pages (regional accessibility observatories, CCN-CERT reports, consultancy surveys) was not searched.

Official monitoring of accessibility statements. Spain monitors RD 1112/2018 through the OAW, on samples. Spain's report to the Commission for 2022–2024 (*Informe sobre el resultado del seguimiento. Periodo 2022-2024*, March 2025) gives 1,065, 1,510 and 1,500 websites in the simplified monitoring of 2022, 2023 and 2024 (364, 750 and 775 of them local), and 64, 62 and 63 websites in the in-depth monitoring. The 2025 simplified monitoring (*Informe global del seguimiento simplificado de sitios web*, 29/10/2025) covered 1,552 websites, 730 of them local, with 20 accessibility checks (text alternatives, headings, lists, contrast, ...), and estimated "Plenamente conforme 0.58%", "Parcialmente conforme 76.69%", "No conforme 22.73%"; none of its checks concerns the statement. The 2025 in-depth monitoring (*Informe agregado del seguimiento en profundidad de sitios web*, January 2026) did check the statement on its 63 websites: "Los requisitos 12.1 Documentación del producto hacen referencia a si se ha proporcionado la Declaración de Accesibilidad y si esta documentación es accesible"; requirement 12.1.1 was judged "Conforme" on 87.3% of them. **So whether a public body provides an accessibility statement is already measured officially, by experts, on a sample of about 63 websites a year, together with what the statement says.** We found no official figure on whether the home page links to the statement, or on how old the statement's preparation or review date is, and no census of all bodies.

Academic studies of statements and of municipal websites. Jonsson et al. (2023) assessed the statements of 37 Swedish public healthcare providers: "All but one of the 37 evaluated healthcare providers published an accessibility statement. None of the healthcare providers fully met the requirements for accessibility statements". Maciejewska (2026) compared declared and measured accessibility on 24 government websites in six member states and found that "Romania is the only country where all four websites lack any WAD-compliant accessibility statement". For Spain, Martín-Herrero and Padilla Castillo (2024) examined "las webs de los 52 ayuntamientos más importantes [de] España" and found that "solo tres cuentan con certificaciones externas de terceros"; Pastor Albaladejo and Sánchez Medero (2023) measured the accessibility of municipal transparency portals with sixteen indicators; Escudero Mancebo and Alvar Herrero (2024) compared the automated accessibility of public and private websites and found that "the public sector has made progress in meeting the regulations". Earlier Spanish studies also work on samples, such as the 62 municipalities of more than 100,000 inhabitants (Sánchez-Labela Martín, Simelio and Moreno-Sardá, 2017). Elsewhere, studies of municipal accessibility use samples of 57 English councils (Wharton et al., 2026) or the 100 smallest Romanian municipalities (Pribeanu, 2026). None of these covers every municipality of a country, and none reads the statement's date.

security.txt and HSTS. Poteat and Li (2021) monitored `security.txt` adoption among top websites for 15 months; Findlay and Abdou (2022) measured it over the Tranco top million; Hilbig et al. (2023) concluded that "the overall adoption of security.txt remains low, especially among less popular websites". Neef et al. (2025) report "25 new security.txt files" among the 40 companies of Germany's DAX between 2023 and 2025. On government websites, Singanamalla et al. (2020) measured HTTPS adoption worldwide, including the long tail, and "observe an overall lower https rate and a steeper dropoff with descending popularity among government sites compared to the commercial websites"; Latta and Keshvadi (2026) found among Canadian government authority websites that "only 59.73% deploy HTTP Strict Transport Security (HSTS)" and "fewer than 2% support standardized vulnerability disclosure via security.txt"; Waheed and Alazab (2026) found that "Only 11.8% of evaluated Maldivian websites" (government and listed companies) send all of HSTS, X-Frame-Options and X-Content-Type-Options. We found no such measurement for Spanish public bodies.

AI crawlers and llms.txt. Longpre et al. (2024) audited the `robots.txt` of 14,000 domains behind AI training corpora and observed "a rapid crescendo of data restrictions from web sources"; Liu et al. (2025) studied whether artists can use `robots.txt` against AI crawlers; Steinacker-Olsztyn et al. (2025) found that "60.0% of reputable sites disallow at least one AI crawler, compared to just 9.1% of misinformation sites". None of them looks at public administrations. The arXiv query for `llms.txt` returned two papers, neither a prevalence measurement.

Tools. GitHub has `security.txt` parsers and validators (e.g. `eikendev/sectxt` , "A library & tool for probing, parsing, and validating security.txt files as specified in RFC 9116"), an HTTPS scanner built for US government domains (`cisagov/pshtt` , "Scan domains and return data based on HTTPS best practices"), a list of AI crawlers to block (`ai-robots-txt/ai.robots.txt`) and a generator of Spanish accessibility statements (`ralcarazm/generador-declaraciones-accesibilidad`). We found no repository that publishes a census of Spanish public-sector websites for these indicators.

What this study adds. Narrowly: (1) a census rather than a sample — every Spanish public body for which a website could be found, with the per-entity table and the code; (2) for accessibility statements, the two things for which we found no official figure: whether the home page links to the statement, and how old its visible date is. Whether a statement exists and whether its content is right is measured better by the OAW's expert sample, and we do not replace it; (3) `security.txt` (with two levels of validity), HSTS and AI-crawler policy for the Spanish public sector, which no study or official monitor we found has measured for Spain; and (4) the measurement of the measurement: how many public bodies have a findable website and how many published URLs no longer lead to the body's site.

3. Legal and technical baseline

Accessibility statement. Royal Decree 1112/2018 (BOE-A-2018-12699), art. 15.1, as consolidated in the BOE (read 2026-10-02), says: "Las entidades responsables de las webs y aplicaciones para móviles proporcionarán una declaración de accesibilidad detallada, exhaustiva y clara sobre la conformidad de sus respectivos sitios web y aplicaciones para dispositivos móviles con lo dispuesto en este real decreto. Dicha declaración será actualizada periódicamente, como mínimo una vez al año, o cada vez que se realice una revisión de accesibilidad conforme a lo especificado en el artículo 17. [...] En el caso de los sitios web, la declaración se publicará en el sitio web correspondiente estando disponible su acceso desde todas las páginas del sitio web con un enlace denominado «Accesibilidad» o su equivalente en el idioma en el que se encuentre disponible la página." The article requires the statement to be updated at least once a year; it does not require the page to show a date. Our date check therefore measures whether a statement *shows* a preparation or review date from the last 365 days. A statement without such a visible date is not proof that it was not updated. The link check follows the last sentence: the home page is one of "todas las páginas".

The BOE's note on art. 15.3 says that Constitutional Court judgment STC 100/2019 (BOE-A-2019-11912) "declara inconstitucional y nulo el inciso destacado del apartado 3 y que el texto restante de dicho apartado invade las competencias autonómicas y carece de carácter de legislación básica". Art. 15.3 concerns who approves the instructions and model for statements; art. 15.1, which we test, is not affected.

security.txt. RFC 9116 (2022) defines `/.well-known/security.txt` . Its requirements that we test are quoted in §6.3. No Spanish rule we found requires the file. The Esquema Nacional de Seguridad, Royal Decree 311/2022 (BOE-A-2022-7191), "es de aplicación a todo el sector público" (art. 2.1); its consolidated

text does not mention `security.txt`. The NIS2 Directive ((EU) 2022/2555) had to be transposed by 17 October 2024. The Council of Ministers approved a draft bill (*anteproyecto de Ley de Coordinación y Gobernanza de la Ciberseguridad*) on 14 January 2025 (La Moncloa, referencia del Consejo de Ministros). A query of the BOE's API for *consolidated legislation* on 2026-10-02, for titles containing *ciberseguridad*, returned nothing newer than 2019; a very recent law could be missing from that API, so we say only that we found no transposing law there. We treat `security.txt` as a voluntary, widely recommended practice and measure it.

HTTPS and HSTS. RFC 6797 defines HSTS. No Spanish rule we found names HTTPS, TLS or HSTS for public websites. The ENS's measure on the confidentiality of communications reads "Se emplearán redes privadas virtuales cifradas cuando la comunicación discurra por redes fuera del propio dominio de seguridad" (Annex II, mp.com.2), and its measure on web services (mp.s.2) does not name HTTPS, TLS or HSTS either.

AI crawlers. There is no obligation. `robots.txt` (RFC 9309) is how a site states its policy; `llms.txt` is an informal proposal (llmstxt.org, 2024) for a Markdown file that guides language models.

4. Data

Entities. The 8,132 municipalities in the Registro de Entidades Locales (Ministerio de Política Territorial; Excel export downloaded 2026-10-02). The export lists two municipalities twice (Alhama de Granada and Soba, rows that differ only in an annotation); we keep one row of each. The result matches, code for code, INE's *Relación de municipios y códigos por comunidades autónomas y provincias a 1 de enero de 2026* (<https://www.ine.es/daco/daco42/codmun/diccionario26.xlsx>), which lists 8,132 municipality codes. Plus 143 curated entities: 19 regional governments (including Ceuta and Melilla), 52 provincial bodies (38 *diputaciones*, 3 *diputaciones forales*, 7 *cabildos*, 4 *consells insulars*), 50 public universities and 22 ministries. Ceuta and Melilla appear both as municipalities and as autonomous cities with one website each; we count each once, as a regional government.

Municipal URLs. We found no official national file listing municipal websites. The REL has no web field; the DIR3 download page rejected automated requests; the BDGEL entity page shows no website; Castilla y León's open register has no web field. Two sources had URLs: Castilla-La Mancha's *Directorio de Entidades Locales* (official, June 2025 edition, 521 URLs joined) and Wikidata, queried by INE municipality code (P772) rather than by class, which recovers cities that a class-based query misses (Barcelona). One URL override was written by the agent (Bilbao's only Wikidata value was its tourism site), and three tourism-only values were dropped.

MUNICIPALITIES	N	WITH URL	%	% OF POPULATION
< 1,000 inhabitants	4,980	3,612	72.5	82.5
1,000–5,000	1,827	1,717	94.0	95.1
5,000–20,000	893	889	99.6	99.5
20,000–100,000	365	363	99.5	99.6
≥ 100,000	67	67	100	100
all	8,132	6,648	81.8	98.9

Coverage is lowest in Navarra, Castilla-La Mancha before the regional file, La Rioja and Castilla y León (`data/coverage.csv`), all regions with many very small municipalities. Some of them have no website of their own, or a page inside a provincial portal that no source records. The 1,484 municipalities without a URL are therefore not a random sample: they are smaller and, plausibly, less resourced. Every rate below describes municipalities *with a findable URL*, and the true rates for the whole set are probably lower.

5. Method

5.1 The scan

The scanner (`scripts/scan.py` , Python standard library) fetched, for each entity, `http://host/robots.txt` (which also tells whether plain HTTP is upgraded to HTTPS), the home page, `https://<final host>/.well-known/security.txt` , and either the accessibility statement or `/llms.txt` . A deterministic one-in-five split of hosts (arm L) always spent the fourth fetch on `llms.txt` , so its prevalence is estimated on a subsample that did not depend on the site. At most one request per second went to a host, redirect hops included, with 48–96 hosts in parallel; the scan ran from 19:08 to 19:20 UTC on 2026-10-02 from one Spanish residential IP, with the User-Agent `EasyByteLab-research/0.1` (contact: `contact@easybyte.es`) . No response body was stored, and no `security.txt` contact value was kept. `METHOD.md` gives the exact rules and the fetch budget, with its exceptions.

5.2 Is the URL the entity's website?

A URL in a register is not always the body's site. We judge every home page that answered 200 with HTML by rules applied in this order (`scripts/aggregate.py` , `site_check`), and list each exclusion with its rule in `data/exclusions.csv` :

1. the URL or its final page is on a video, social-network or portal platform (YouTube, Facebook, Yahoo, ...) — 2 municipalities;
2. it is a domain marketplace or a parked/for-sale page (HugeDomains, DropCatch, "Dominio en reserva", ...) — 18;
3. it is a postcode or place directory — 2;

4. it is a hosting-panel login (Plesk's `/login_up.php` , among them the URL given for Vigo) or a server default page ("Index of /", "Domain Default page", "Coming Soon", ...) — 48;
5. it redirected to another registrable domain whose page and URL carry no form of the municipality's name (betting and other spam on expired domains, a provincial portal's home page) — 30;
6. its title names *another* municipality's council, or it names a council or public body but neither it nor its URL names this municipality — 37 (36 of them domains of small municipalities in Salamanca that served the site of the council of Alba de Tormes);
7. it has text and a title, but neither the municipality's name nor any council wording (mostly hijacked domains) — 42.

Rules 5–7 apply only to municipalities. They use the scan's page and, for the 736 municipal home pages whose identity or accessibility link needed a second look (name not found, near-empty page, other domain, panel path, widget link), a re-fetch made the same evening (§5.4; 617 of them answered). In all, 179 municipal URLs were excluded (161 from Wikidata, 18 from the Castilla-La Mancha directory). Of the 5,121 municipal home pages kept, 5,035 show the municipality's name, 8 show council wording and carry the name in their URL, and 78 could not be confirmed either way (mostly pages built by JavaScript, or servers that did not answer the re-fetch); they are kept and counted.

5.3 robots.txt: what went wrong, and what we deleted

The scanner was meant to honour `robots.txt` for every page it fetched. It did not, for three reasons, found by the adversarial review (§11): it ignored any `robots.txt` served with an HTML Content-Type, treating it as absent; it never checked `robots.txt` before requesting `/.well-known/security.txt` (an earlier draft claimed that RFC 9116 exempts that file from `robots.txt` ; it does not, and RFC 9116 does not mention `robots.txt`); and it did not read the `robots.txt` of hosts it was redirected to. RFC 9309 is clear on the first point: "If the crawler successfully downloads the robots.txt file, the crawler MUST follow the parseable rules" (§2.3.1.1), whatever its Content-Type.

To find every request that the `robots.txt` in force did not allow, we re-read the `robots.txt` of all 8,055 hosts the scan or the pilot had sent a request to, between 21:45 and 21:58 UTC the same day, and re-decided every recorded request with an RFC 9309 parser (`scripts/robots9309.py` ; longest match, `*` and `$` wildcards, groups merged) for the product token the scan had sent. Where the scan itself had seen a 5xx or no answer, the RFC's "MUST assume complete disallow" (§2.3.1.4) applied. A record is kept only if the request was allowed; when the `robots.txt` could not be re-read, the record is discarded too.

- **168 entities lost their whole record** (home page, statement, `security.txt` , `llms.txt`): 123 whose `robots.txt` did not allow the home page or a redirect hop of it — 100 of them municipalities in the province of Castelló/Castellón whose shared platform serves `User-agent: * / Disallow: / as text/html` — and 45 whose `robots.txt` could not be re-read to verify. Only their `robots.txt` -derived fields remain. They are reported as *not measurable: robots.txt disallows* (or *not verifiable*) and stay in the denominator of reachability.
- For other entities, single requests were discarded: `security.txt` (88), `llms.txt` (85), the statement (4). In all, 845 entities had at least one request that their `robots.txt` did not allow, and 921 had at least one record discarded.
- A second pass had fetched `security.txt` for 672 entities whose home page `robots.txt` had forbidden or whose host budget was spent; 661 of those fetches were not allowed (or not verifiable) and were discarded.

The discarded data were deleted from our working files, not only left out of the figures; the pilot's published results (`pilot/muni_res.json`) lost the fields it had fetched against two hosts' `robots.txt` . The corrected scanner reads `robots.txt` before every request to a host, redirect hops included, whatever its Content-Type (`METHOD.md` §7). The same re-read, read with the RFC 9309 parser, is the source of every AI-crawler figure in §6.5. For the 4,748 entities whose `robots.txt` the scan had read as plain text and the re-read found again, the scan's parser (Python's `urllib.robotparser`) and the RFC 9309 parser agree on all four headline tokens for 4,743.

5.4 Re-checks made the same evening

Besides the `robots.txt` re-read, and always within what `robots.txt` allows: 758 home pages were fetched again to apply the rules in §5.2 and to re-detect the accessibility link where the first detector had picked a widget's credit link; 116 statements whose first-version date the current extractor rejected were read again in full (§7); and `/.well-known/security.txt` was fetched again on the 360 hosts where the scan had found a file or a "soft 404", to apply the stricter definition below. Each host received at most four requests for resources in this session, `robots.txt` included, with one exception (one host received five, `METHOD.md` §6).

6. Results

6.1 Reachability

Of 6,791 entities with a URL, 61 could not be measured because of the scanner's own fetch budget or because they are the same website as another entity (59 and 2). Of the remaining 6,730, 5,245 (77.9%) returned their own home page. The 1,467 municipalities that did not (of 6,588 measured) break down as follows: the domain no longer resolves (244), the server returned an HTTP error (200), *not measurable: robots.txt disallows* (212, of which 123 are the records deleted in §5.3), `robots.txt` answered 5xx so we did not crawl (193), the URL does not lead to the municipality's site (179, §5.2), the server did not answer (103), the server answered HTTP 202 with no content, typically a bot challenge (101; 63 of them in the province of Málaga), the server refused our client (94; HTTP 401, 403 or 429), a redirect loop that our cookie-less client could not leave (86, all on one *sede electrónica* platform), *not verifiable* (44, §5.3) and a 200 response that is not HTML (11). Four of the 22 ministries' home pages answered 403 to our client; some of these refusals depend on the User-Agent (`README.md`).

6.2 HTTPS and HSTS

	REACHABLE	SERVED OVER HTTPS ¹	CERTIFICATE VERIFIES ²	HTTP → HTTPS ³	HSTS	HSTS ≥ 1 YEAR
municipalities	5,121	97.8%	93.9%	77.4%	32.3%	24.9%
provincial councils	44	100%	95.5%	90.2%	56.8%	52.3%
universities	46	100%	100%	85.4%	54.3%	37.0%

	REACHABLE	SERVED OVER HTTPS ¹	CERTIFICATE VERIFIES ²	HTTP → HTTPS ³	HSTS	HSTS ≥ 1 YEAR
regional governments	16	100%	100%	73.3%	43.8%	43.8%
ministries	18	100%	83.3%	93.3%	77.8%	61.1%
all	5,245	97.8%	94.0%	77.6%	32.9%	25.4%

Denominators: "served over HTTPS" over reachable home pages; certificate and HSTS columns over home pages served over HTTPS (5,130 in all); HTTP → HTTPS over reachable entities whose plain-HTTP `robots.txt` request got an answer (5,201). Per-cell counts are in `data/summary_by_type.csv`.

¹ The scanner asks for `https://` first. This column says that the site answers over HTTPS when asked, not that visitors are moved to it. ² Verified by OpenSSL 3.6.2 (Python 3.14.4) with the Mozilla roots and no intermediate-certificate fetching. A server that omits its intermediate certificate fails here although most browsers repair the chain, so this column measures strict TLS hygiene, not what a visitor sees. ³ Whether `http://host/robots.txt` ended on an `https://` URL: a proxy for an HTTP-to-HTTPS redirect, measured on that one request.

HTTPS is nearly universal: 5,130 of 5,245 reachable home pages (97.8%). HSTS is the exception: 1,688 of the 5,130 (32.9%) send it. 40 of those send it over a connection whose certificate did not verify; RFC 6797 §8.1 has the browser note an HSTS host only when "there are no underlying secure transport errors or warnings", so HSTS that a browser would apply is 1,648 of 5,130 (32.1%). HSTS varies more by region than by size: 299 of 395 municipal home pages served over HTTPS in the Valencian Community (75.7%) and 42 of 57 in the Balearic Islands (73.7%) send it, against 3 of 63 in Asturias (4.8%), 4 of 80 in Navarra (5.0%) and 23 of 412 in Aragón (5.6%). Municipal sites of one region often share a provider, and the header is usually a server setting; we did not identify the providers, so this is an inference from the clustering, not a finding.

6.3 security.txt

`/.well-known/security.txt` got an answer from the servers of 5,570 entities (requests that `robots.txt` did not allow are not counted). We report two levels, for the 12 entities (11 files) that serve a file with at least a `Contact` or an `Expires` field:

- **Has the required fields:** served over HTTPS, at least one `Contact` field, exactly one `Expires` field, whose value is a date-time that has not passed. RFC 9116 says of `Contact`: "This field MUST always be present in a "security.txt" file" (§2.5.3), and of `Expires`: "This field MUST always be present and MUST NOT appear more than once" (§2.5.5); a past `Expires` marks data that "is considered stale and should not be used". **8 of 5,570 entities (7 files):** the municipalities of Madrid, Málaga, Adeje, Fuenllana, Bargas and Pobladura del Valle, and the Deputación da Coruña, whose file also serves the municipality of Vilasantar on the same host.
- **Strictly valid under RFC 9116:** the above, and every other requirement we can test: "It MUST have a Content-Type of "text/plain" with the default charset parameter set to "utf-8"" (§3); every `Contact` value is a URI ("The value MUST follow the URI syntax described in Section 3 of [RFC3986]. This

means that "mailto" and "tel" URI schemes must be used when specifying email addresses and telephone numbers", §2.5.3); web URIs begin with `https://` ; `Expires` follows RFC 3339 (seconds included; lowercase `t` and `z` accepted, as in the RFC's own example `2021-12-31T18:37:07z`); `Preferred-Languages` appears at most once; every line is a comment or a field; the body is UTF-8. We do not verify OpenPGP signatures. **4 of 5,570 entities:** Madrid, Málaga, Adeje and Fuenllana. The Deputación da Coruña, Pobladura del Valle and Bargas send `text/plain` without the `charset` parameter, and Bargas's `Contact` is an e-mail address without `mailto:` .

Of the other four files, two have expired (Castroverde, Pezuela de las Torres) and two lack the mandatory `Expires` field (Junta de Andalucía, Universitat Politècnica de València). No ministry publishes one. These levels were applied on the evening re-fetch of each file.

The more frequent outcome is a trap for naive checkers: **329 of the 5,570 (5.9%) answer the path with HTTP 200 and something that is not a security.txt file** — mostly an HTML page (a home page, a search page or a "not found" page with the wrong status), sometimes a text without any security.txt field. A tool that only tests for "200 OK" would report 27 times more `security.txt` files than exist.

6.4 Accessibility statement

A link to the accessibility statement is visible in the static HTML of **2,492 of 5,245 reachable home pages (47.5%)**: 16 of 18 ministries (88.9%), 37 of 44 provincial councils (84.1%), 38 of 46 universities (82.6%), 13 of 16 regional governments (81.2%) and 2,388 of 5,121 municipalities (46.6%). Among municipalities it rises with size (1,219 of 2,614 under 1,000 inhabitants, 46.6%; 194 of 319 at 20,000–100,000, 60.8%; 48 of 57 above 100,000, 84.2%) and varies strongly by region (59 of 65 in Asturias, 90.8%; 162 of 217 in Extremadura, 74.7%; 633 of 862 in Catalonia, 73.4%; 5 of 113 in La Rioja, 4.4%; 20 of 146 in the Basque Country, 13.7%; 31 of 163 in the Community of Madrid, 19.0%). The detector ignores links to accessibility widget and plug-in vendors (the first version counted 67 credit links such as "Accessibility by WAH" as statement links; those 67 home pages were re-read). This figure is neither a lower nor an upper bound: it misses links injected by JavaScript and links labelled only with an icon, and it may still count a few links that lead elsewhere.

Of 2,012 linked pages requested, 1,914 (95.1%) answered 200; 448 of the HTML pages among them lack statement wording ("declaración", "RD 1112/2018" or equivalents) and are left out of the date figures, as are PDFs.

Dates. Of **1,247 statement pages** read with the audited extractor (§7), **908 (72.8%) show a preparation or review date, and 160 (12.8%; 95% CI 11.1–14.8) show one from the last 365 days.** Counting every linked page, statement wording or not, the figures are 909 and 160 of 1,629 (55.8% and 9.8%). The dates are old: of the 908, 35 are from 2018–2021, 203 from 2022, 97 from 2023, 340 from 2024, 110 from 2025 and 123 from 2026. As said in §3, an old or missing date on the page is not proof that the statement was not updated.

The six most frequent dates account for 479 of the 908 dated statements (52.8%), and each occurs in a single province: 28 June 2022 on 148 statements in Burgos, 27 March 2024 on 140 in Valencia, 6 September 2023 on 57 in Granada, 18 February 2026 on 50 in Badajoz, 26 April 2024 on 48 in Albacete and 11 December 2023 on 36 in Jaén. Identical dates across dozens of municipalities of one province are consistent with statements produced in bulk by a provincial service or a common provider; we did not identify who produces them. The same clustering shows in the share of fresh statements by region: 50 of 50 in Extremadura and 89 of 424 in Catalonia (21.0%), against 0 of 221 in Andalusia (none shows a date

from the last year), 1 of 163 in Castilla y León and 1 of 148 in the Valencian Community. Under art. 15.1 the obligation stays with each responsible body, whoever writes the text.

Statements of large municipalities and of all non-municipal entities are not in these figures. The first 586 entities scanned (all 143 non-municipal entities and the largest municipalities) were read by the first version of the extractor before the scanner stored the sentence each date came from, so their dates cannot be checked against the audited version; all 287 statements fetched for them are left out of the date figures, dated or not. That is why `data/summary_by_size.csv` has almost no date data above 20,000 inhabitants.

6.5 AI crawlers and llms.txt

These figures come from the evening re-read of `robots.txt` (§5.3), parsed under RFC 9309, over the 6,012 entities whose `robots.txt` answered (a file, or a 4xx meaning no restrictions) and whose URL leads to their own site or was not judged otherwise. **326 of the 6,012 (5.4%) disallow `/` to at least one of GPTBot, ClaudeBot, Google-Extended or CCBot**; 165 (2.7%) do so by naming the crawler, 194 (3.2%) block all four, and 171 (2.8%) disallow `/` to every crawler not named in the file (`User-agent: *` , which may still allow named search engines). By token, over the same 6,012: Bytespider 345 (5.7%), GPTBot 312 (5.2%), meta-externalagent 297 (4.9%), ChatGPT-User 294 (4.9%), CCBot 260 (4.3%), ClaudeBot 257 (4.3%), OAI-SearchBot 230 (3.8%), Google-Extended 200 (3.3%), PerplexityBot 185 (3.1%).

Blocking is clustered by region: 52 of 70 municipalities in Asturias (74.3%, all by naming the crawlers), 55 of 187 in the Basque Country (29.4%; 50 by naming), 103 of 512 in the Valencian Community (20.1%; 100 through `User-agent: *` / `Disallow: /` , the Castelló platform of §5.3), 13 of 84 in the Canary Islands (15.5%) and 9 of 66 in the Balearic Islands (13.6%), against no more than 7.1% (3 of 42, Murcia) in any other region. As with HSTS, this pattern fits shared templates better than decisions taken site by site, but we did not identify the providers. Higher tiers block more often: 9 of 48 provincial councils (18.8%), 9 of 50 universities (18.0%) and 2 of 18 regional governments (11.1%), but 1 of 22 ministries.

A spec-like `llms.txt` (200, not HTML, starting with a Markdown heading) appears on **7 of 1,020 reachable sites in the random arm (0.7%; 95% CI 0.3–1.4)**, and on 26 sites across both arms, including Madrid and Gijón. Another 66 sites answer `/llms.txt` with a non-HTML 200 that is not an `llms.txt` .

7. How reliable is the one non-trivial check?

Reading a date from a statement is the only step that interprets text. The first extractor (v1) accepted a date after generic words such as *fecha*, *data* or *última*. Looking at its output during the scan, the agent found non-statement dates: a Catalan municipal platform prints "Darrera actualització: 23.05.2024 | 13:36" (the page's last edit) in every footer, and one page showed the current date and time. No record of that first look was kept, so we give no rate for it. The extractor was replaced mid-scan with v2, which only accepts dates after words that refer to preparing, reviewing or updating the statement (*preparada*, *revisión*, *revisada*, *actualizada*, *elaborada*, *berrikuspena*, *eguneratu*, *updated on*, ...) and ignores the two footer patterns.

The reproducible evidence is this. For entities scanned under v1 after the scanner began storing the sentence each date came from, v2 was run on that sentence: 345 of 461 v1 dates were confirmed and 116 rejected (25.2%). Those 116 statements were then read again in full with v2 the same evening: 94 do show

a v2 date elsewhere on the page (v1 had picked a footer date over it), and those readings are the ones used. The sentences are in `data/records.jsonl.gz`, so anyone can repeat the check.

Precision of v2: the agent audited 40 dated statements drawn at random from the 908 (`scripts/draw_audit_sample.py`, seed 8), reading the stored sentence; the verdicts and a note for each are in `data/date_audit_v2.csv`. All 40 dates are a preparation, review or update date of the statement and the latest one in the sentence (95% CI for precision 91.2–100). Two of the 40 (an Andalusian template) give the date of the site's accessibility review that the statement reports, which we count as a review date. The independent reviewer separately re-fetched 25 other dated statements and found 25 correct. Recall is not measured: a date written in an unusual form, inside a PDF, in an image, or after a word v2 does not know (for example Catalan *actualització*, which v2 excludes because of the footer above) is missed, so "shows a date" is a lower bound and "shows a date from the last year" is too.

8. Limitations

- **Coverage bias.** 1,484 municipalities (18.2%), almost all small, have no URL in any source we found. Wikidata is volunteer-maintained; 161 of its 6,126 municipal URLs (and 18 of the 521 from the Castilla-La Mancha directory) led to something that is not the municipality's site, and 211 (and 33 of the directory's) to domains that no longer resolve.
- **Site identity.** The rules of §5.2 are heuristics. 78 reachable municipal home pages could not be confirmed as the council's and are kept; some excluded pages may be a council's real but unusual site.
- **Curated lists.** The URLs of the 143 non-municipal entities were written by the agent from general knowledge and are validated only by the scan; the reviewer fetched a sample of them and all 22 ministries, and every page that answered belonged to the named body (§11).
- **One day, one vantage point, two moments.** The scan ran from a Spanish residential IP at 19:08–19:20 UTC; the re-checks ran 2½ hours later from the same IP. A site that changed in between is measured partly at each time. WAFs that block unknown clients count as not reachable (94 municipalities refused, 101 answered with a 202 challenge, four ministries); some of these refusals depend on the User-Agent.
- **Robots.txt read later than used.** For the hosts whose `robots.txt` the scan had read as plain text or not at all, the decision to keep a record rests on the evening re-read. A file that changed in between could make us keep or drop a record wrongly; when the re-read failed we dropped the record.
- **Static HTML.** Links and dates rendered by JavaScript are not seen. The first scanner also read as binary the gzip bodies that a few servers send without being asked; the re-checks met this on two municipal home pages (one was then read correctly), and the corrected scanner decodes them.
- **Budget accounting.** See `METHOD.md` §2: the limit of four resource fetches per host had exceptions in the first scan.
- **Not measured.** Accessibility itself, statement content, security contacts, *sedes electrónicas* as such, municipal companies.

9. A tool any administration could run on its own site

The checks in this study are cheap enough to run continuously, and the obligation in art. 15.1 ("desde todas las páginas") is one that only the site owner can verify exhaustively, since an outside census should

not crawl every page. We propose — and have not built — `gov-web-lint`, a free command-line tool and CI step that an administration points at its own site. It would crawl the site within limits the owner sets, and report: pages without an "Accesibilidad" link; the statement's last review date and the days left before it turns a year old; whether `security.txt` is valid under RFC 9116 and how long until it expires; HTTPS, HSTS and certificate-chain completeness; and the site's effective policy for each AI crawler. A body that hosts many municipal sites could run it once for all of them.

10. Tool and data

`run.sh` rebuilds every table in `data/` from the scan records and checks the figures quoted in this paper and in `README.md` against `data/summary.json` (`scripts/check_numbers.py`). `data/records.jsonl.gz` holds the cleaned per-entity scan records (no response bodies, no `security.txt` contacts, no page titles; e-mail addresses and phone numbers masked in the stored statement sentences).

`data/entities.csv` has one row per entity (8,275) with every indicator; `data/exclusions.csv` lists the URLs judged not to be the entity's site, with the rule; the `summary_by_*.csv` files have counts, denominators and percentages (blank when the denominator is under 10); `data/coverage.csv` has URL coverage; `data/build_meta.json` has the counts of the `robots.txt` re-read and the re-checks; `data/date_audit_v2.csv` has the audit.

11. Automation and review

This study was run by AI agents working in a terminal under EasyByte's supervision:

- **Done by an AI agent:** searching for prior work (arXiv, Crossref and Semantic Scholar APIs, GitHub, the official monitoring reports; §2); searching for and testing URL sources (REL, DIR3, BDGEL, regional open-data portals, Wikidata); writing the curated list of non-municipal entities from general knowledge; writing the scanner, the `robots.txt` parser, the population builder, the record builder and the aggregator; running the scan and the re-checks; inspecting the date extractor and replacing it mid-scan; auditing 40 dates; designing the rules that decide whether a URL is the entity's site; reading the legal texts in the BOE and the RFCs and quoting them; writing this paper, `METHOD.md` and `README.md`.
- **Checked by an independent AI reviewer:** a second agent, asked to be adversarial, recomputed the figures from the data, re-fetched a sample of sites, statements and `robots.txt` files, and verified each quotation against its source. It found the `robots.txt` failures of §5.3, the non-council URLs of §5.2, a denominator error in the date figures, the duplicated municipalities and a dozen smaller errors. Every blocker and correction it listed was addressed in this version; the reviewer has not yet re-read the corrected version.
- **Checks:** wherever a document says something was checked (the findings, the code, the curated URLs, the choice of indicators and denominators, the legal quotations, the per-entity table), an AI agent checked it.

No administration was contacted, and nothing was submitted to any form.

Competing interests

EasyByte, the cooperative behind EasyxLab, is a software company that could offer the kind of tool proposed in §9; that tool does not exist today. EasyByte has no contract with any of the bodies measured.

References

- Jonsson, M., Gustavsson, C., Gulliksen, J., Johansson, S. (2023). How have public healthcare providers in Sweden conformed to the European Union's Web Accessibility Directive regarding accessibility statements on their websites? Universal Access in the Information Society. <https://doi.org/10.1007/s10209-023-01063-1>
- Maciejewska, K. (2026). Accessibility Declarations vs. Automated Audit Results: A Cross-Country Study of Government Websites in Six EU Member States. Communications of International Proceedings. <https://doi.org/10.5171/2026.4731426>
- Martín-Herrero, J. M., Padilla Castillo, G. (2024). Hacia un enfoque de ciudades inteligentes inclusivas: la accesibilidad web de los principales ayuntamientos de España. Revista Española de Desarrollo y Cooperación. <https://doi.org/10.5209/redc.92873>
- Sánchez-Labela Martín, I., Simelio, N., Moreno-Sardá, A. (2017). El acceso web para personas con capacidades limitadas en los ayuntamientos españoles. Cuadernos.info. <https://doi.org/10.7764/CDI.41.1061>
- Pastor Albaladejo, G. M., Sánchez Medero, G. (2023). Transparencia activa inclusiva en los Ayuntamientos españoles. Revista Española de la Transparencia. <https://doi.org/10.51915/ret.265>
- Escudero Mancebo, D., Alvar Herrero, J. (2024). Assessment of web accessibility regulation adherence by institutions with compliance obligations. DYNA. <https://doi.org/10.52152/d11128>
- Wharton, R., Kavanagh-Smith, L., Douce, C. (2026). Evaluating the accessibility of local authority public facing websites in England. Universal Access in the Information Society. <https://doi.org/10.1007/s10209-025-01269-5>
- Pribeanu, C. (2026). Web accessibility of small Romanian municipalities. Annals of the Academy of Romanian Scientists, Series on Economy, Law and Sociology. <https://doi.org/10.56082/annalsarscieco.2026.2.19>
- Poteat, T., Li, F. (2021). Who you gonna call? An empirical evaluation of website security.txt deployment. ACM Internet Measurement Conference. <https://doi.org/10.1145/3487552.3487841>
- Findlay, W., Abdou, A. (2022). Characterizing the Adoption of Security.txt Files and their Applications to Vulnerability Notification. Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb). <https://doi.org/10.14722/madweb.2022.23014>
- Hilbig, T., Geras, T., Kupris, E., Schreck, T. (2023). security.txt Revisited: Analysis of Prevalence and Conformity in 2022. Digital Threats: Research and Practice. <https://doi.org/10.1145/3609234>
- Neef, S., Schlunke, C., Hennig, A. (2025). Do (Not) Tell Me About My Insecurities: Assessing the Status Quo of Coordinated Vulnerability Disclosure in Germany Amid New EU Cybersecurity Regulations. European Symposium on Usable Security. <https://doi.org/10.1109/EuroUSEC69254.2025.00020>; arXiv:2606.25950.

- Singanamalla, S., Jang, E., Anderson, R. J., Kohno, T., Heimerl, K. (2020). Accept the Risk and Continue: Measuring the Long Tail of Government https Adoption. ACM Internet Measurement Conference. <https://doi.org/10.1145/3419394.3423645>
- Latta, I., Keshvadi, S. (2026). Evaluating the Security Posture of Canadian Government Authority Websites. Canadian Conference on Electrical and Computer Engineering. <https://doi.org/10.1109/CCECE68150.2026.11609975>
- Waheed, L., Alazab, A. (2026). Measuring Baseline Web Security Posture: Tier-1 HTTP Security Header Adoption in the Maldives and the WSHS-B Governance Metric. Future Internet. <https://doi.org/10.3390/fi18060313>
- Longpre, S., Mahari, R., Lee, A., Lund, C., Oderinwale, H., et al. (2024). Consent in Crisis: The Rapid Decline of the AI Data Commons. arXiv:2407.14933.
- Liu, E., Luo, E., Shan, S., Voelker, G. M., Zhao, B. Y., Savage, S. (2025). Somesite I Used To Crawl: Awareness, Agency and Efficacy in Protecting Content Creators From AI Crawlers. ACM Internet Measurement Conference. <https://doi.org/10.1145/3730567.3732913>
- Steinacker-Olsztyn, N., Gosain, D., Dao, H. (2025). Is Misinformation More Open? A Study of robots.txt Gatekeeping on the Web. arXiv:2510.10315.
- Observatorio de Accesibilidad Web (2025). Informe sobre el resultado del seguimiento. Periodo 2022-2024 (March 2025), <https://administracionelectronica.gob.es/PAe/accesibilidad/2024InformeSeguimientoPDF> (read 2026-10-02).
- Observatorio de Accesibilidad Web (2025). Informe global del seguimiento simplificado de sitios web (29/10/2025), and (2026) Informe agregado del seguimiento en profundidad de sitios web (January 2026), linked from https://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Accesibilidad/Informes-Resultados-Seguimiento-Anuales/Resultados-Seguimiento-2025.html (read 2026-10-02).
- Real Decreto 1112/2018, de 7 de septiembre, sobre accesibilidad de los sitios web y aplicaciones para dispositivos móviles del sector público. BOE-A-2018-12699, consolidated text, <https://www.boe.es/buscar/act.php?id=BOE-A-2018-12699> (read 2026-10-02).
- Tribunal Constitucional, Sentencia 100/2019, de 18 de julio. BOE-A-2019-11912.
- Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad. BOE-A-2022-7191, <https://www.boe.es/buscar/act.php?id=BOE-A-2022-7191> (read 2026-10-02).
- Directive (EU) 2022/2555 (NIS2), art. 41.
- La Moncloa, Referencia del Consejo de Ministros de 14 de enero de 2025, <https://www.lamoncloa.gob.es/consejodeminstros/referencias/Paginas/2025/20250114-referencia-rueda-de-prensa-ministros.aspx>.
- BOE, open-data API for consolidated legislation, <https://www.boe.es/datosabiertos/api/legislacion-consolidada> (query of 2026-10-02).
- E. Foudil, Y. Shafranovich, "A File Format to Aid in Security Vulnerability Disclosure", RFC 9116, 2022, <https://www.rfc-editor.org/rfc/rfc9116>.
- M. Koster, G. Illyes, H. Zeller, L. Sassman, "Robots Exclusion Protocol", RFC 9309, 2022, <https://www.rfc-editor.org/rfc/rfc9309>.
- G. Klyne, C. Newman, "Date and Time on the Internet: Timestamps", RFC 3339, 2002.

- J. Hodges, C. Jackson, A. Barth, "HTTP Strict Transport Security (HSTS)", RFC 6797, 2012.
- J. Howard, "The /llms.txt file", <https://llmstxt.org/> (2024).
- Ministerio de Política Territorial, Registro de Entidades Locales, export of municipalities, <https://registroentidadeslocales.mpt.es/> (2026-10-02).
- Instituto Nacional de Estadística, Relación de municipios y códigos por comunidades autónomas y provincias a 1 de enero de 2026, <https://www.ine.es/daco/daco42/codmun/diccionario26.xlsx> (2026-10-02).
- Junta de Comunidades de Castilla-La Mancha, Directorio de Entidades Locales (June 2025), <https://datosabiertos.castillalamancha.es/> (2026-10-02). Fuente de datos: Junta de Comunidades de Castilla-La Mancha.
- Wikidata, properties P772 (INE municipality code), P856 (official website), P576, query of 2026-10-02.

Cite as: EasyxLab (2026). What can a machine verify on Spain's public-sector websites? Study S8. EasyByte Hub S. Coop. Mad. <https://github.com/easybytehub/easyxlab>

Cite this study

CITATION

EasyxLab (2026). What can a machine verify on Spain's public-sector websites? Study S8. EasyByte Hub S. Coop. Mad.
<https://github.com/easybytehub/easyxlab/tree/main/studies/s8-spanish-public-sector-web>

BIBTEX

@techreport{easyxlab_s8,
 title = {What can a machine verify on Spain's public-sector websites?},
 author = {{EasyxLab}},
 institution = {EasyByte Hub S. Coop. Mad.},
 number = {S8},
 year = {2026},
 url = {https://github.com/easybytehub/easyxlab/tree/main/studies/s8-spanish-public-sector-web}
}