

# Are MiCA crypto-asset white papers published in the machine-readable format required since 23 December 2025?

EasyxLab · EasyByte Hub S. Coop. Mad. · Study S15

 Working draft Published 2026-10-03

---

## Abstract

Since 23 December 2025, crypto-asset white papers under MiCA must be drawn up as a single XHTML file with Inline XBRL 1.1 tags for the fields of the ITS Annex; ESMA published the matching taxonomy files on 5 August 2025. We followed the URL that ESMA's interim register gives for every white paper notified or updated since that date and validated what we found. **143 of 440 register rows (32.5%; 95% CI 28.3–37.0%)** lead to an Inline XBRL 1.1 file in ESMA's taxonomy at the registered URL or one link away, which correspond to 141 distinct files; **110 of 143 rows** point to a file that validates with zero errors against ESMA's package, formula assertions included (108 of 141 distinct files). Where they can be compared, the files agree with the register: LEI in 103 of 103 rows, DTI in 134 of 136. One data provider, the registered offeror on those rows, serves 77 of 110 valid rows. Under the access and redirect rules frozen before collection the figure is 139 of 440 (31.6%; 95% CI 27.4–36.1%). Older register entries reach the format in 20 of 634 rows.

---

## 1. Introduction

MiCA (Regulation (EU) 2023/1114) obliges offerors and persons seeking admission to trading of crypto-assets to draw up, notify and publish a white paper, and says that it "shall be made available in a machine-readable format" (Art. 6(10); likewise Art. 19(9) and 51(9)). Implementing Regulation (EU) 2024/2984 (the ITS) chose XHTML with Inline XBRL 1.1, applicable from 23 December 2025; ESMA published the taxonomy files, with formula assertions, on 5 August 2025.

The format is useful only if the machine-readable file is where people and machines look for it. ESMA's interim register lists, for each notified white paper, the URL "where it can be possible to retrieve the white paper document". We asked: nine months after the format became applicable, how many white papers notified or updated since then can be retrieved from that URL in the ITS format, and do those files validate and agree with the register?

---

## 2. Prior work

**Search.** On 2026-10-03 we queried OpenAlex, Crossref and the arXiv API (all fields, exact phrase) and GitHub repository search with six queries: "MiCA white paper XBRL", "crypto-asset white paper machine-readable", "MiCA white paper iXBRL", "crypto-asset white papers MiCA empirical", "MiCAR white paper register ESMA" and "Markets in Crypto-Assets white paper disclosure analysis". Crossref answered one

query and OpenAlex another with HTTP 429; the arXiv queries returned nothing. We read the README files of the relevant repositories, and the independent reviewer read the full text of the closest paper. We also read all 51 MiCA Q&As in ESMA's Q&A tool (§3).

### Closest results.

- *micar-register-observatory* (GitHub, 2026-07-07) diffs the register weekly and, for its snapshot of 2026-09-28, classifies white-paper links "by link shape only" (641 landing pages or bare domains, 263 PDF, 156 "XHTML / HTML"), noting that "a format is a deep-lint candidate, not a verified fact, until the document is fetched".
- *micar-whitepaper-linter* (GitHub) runs Annex I **content** checks on drafts, with a pilot on a convenience sample of Title II white papers (July 2026); it does not check Inline XBRL or the ESMA taxonomy.
- *Asset-Class Specific Sustainability Disclosure: Lessons Learned from the EU MiCA Regulation* (arXiv 2609.26932,
  1. analyses MiCA's sustainability indicators and summarises the register "as of 3rd July 2026" (white papers by competent authority); the full text gives no figure on the white-paper format. It sources its register figures from CCRI's MiCA-Monitor ( [mica-monitor.com](https://mica-monitor.com) ), a JavaScript application whose public page shows no format figures without running it; we did not go further.
- *Legal NLP Meets MiCAR* (NLLP 2023) analyses white-paper text before the ITS existed; *WhitePaperXBRL* (GitHub, 2026-01-27) converts PDFs into iXBRL. Neither measures publication.

**What is new.** No published count of registered white papers available in the ITS format, no validation of the published files against ESMA's assertions and no identifier cross-check with the register and GLEIF; this study gives the three for a dated cohort, plus a reusable checker.

---

## 3. Legal and technical baseline

All quotations come from texts downloaded on 2026-10-03 from the Publications Office (Cellar), ESMA's website and ESMA's Q&A tool, and were checked by `grep` on the downloaded text.

- **MiCA, consolidated version in force** (CELEX 02023R1114-20240109, the latest consolidation; its only amendment, Regulation (EU) 2023/2869, adds Art. 110a on the European single access point "from 10 January 2030"). Art. 6(10): "The crypto-asset white paper shall be made available in a machine-readable format." Art. 6(11): "ESMA, in cooperation with EBA, shall develop draft implementing technical standards to establish standard forms, formats and templates for the purposes of paragraph 10." Art. 19(10) and 51(10) say the same for asset-referenced and e-money tokens, and the ITS is adopted "in particular" under "Article 6(11), third subparagraph, Article 19(10), third subparagraph, and Article 51(10) third subparagraph". Art. 8(7): "ESMA shall make the crypto-asset white paper available in the register, under Article 109(2), by the starting date of the offer to the public or admission to trading." Art. 9(1): offerors and persons seeking admission to trading "shall publish their crypto-asset white papers ... on their website, which shall be publicly accessible".
- **ITS 2024/2984** (CELEX 32024R2984; Cellar lists no consolidated version, i.e. no amendment). Art. 2(1): "Crypto-asset white papers shall be drawn up in XHTML format marking the fields set out in the Annex using Inline XBRL 1.1 specifications ... (a) the Inline XBRL instance document containing the crypto-asset white paper shall be submitted as a single XHTML file"; (b) and (c) require the ISO 17442 LEI "where available". Art. 2(2): "Taxonomy shall be used for the drawing up of a crypto-asset white

paper and the elements shall be those set out in Table 2, Table 3 or Table 4 of the Annex." Art. 3: "ESMA may publish machine-readable and downloadable XBRL taxonomy files". Art. 4: "It shall apply from 23 December 2025." Recital (6): white papers "should be human readable and easily accessible without specialised software. The use of Inline XBRL technology ... enables such documents to be at the same time machine-readable and human readable." ESMA's statement of 28 November 2025 (ESMA75-1303207761-6284) says that the taxonomy "to be used to comply with these requirements" was published on 5 August 2025.

- **ESMA Q&As.** Q&A 2845 (answered 12-05-2026): a modified white paper notified after 23 December 2025 "should" be prepared "in the machine-readable format" even if the original predates the requirement, but "the entry into application of the ITS does not constitute a 'significant new factor': older white papers need not be redone. Q&A 2654 (Title II tokens admitted before 30 December 2024, Art. 143(2)) only recalls in passing the Art. 66(3) hyperlink duty. None of the 51 MiCA Q&As (one, on DLT market infrastructures, is unanswered), nor the reporting manual, says whether the version **published on the website** must be the Inline XBRL file.
- **The register** was still five CSV files on 2026-10-03 ("Last update: 30 September 2026"), although ESMA had announced integration into its IT systems for "mid-2026". It lists URLs only; none points to an ESMA-hosted file.

**Consequence for the framing.** Art. 6(11) is the strongest textual link: the ITS format exists "for the purposes of" the duty to make the white paper "available in a machine-readable format", and Art. 9(1) says where white papers are made available to the public. Read with recital (6), this suggests that the copy people find on the website is meant to be the machine- and human-readable file. But no provision or Q&A says in terms that the website copy must be the Inline XBRL file rather than a rendering of a file notified to the authority, and the ITS speaks of the file being "submitted". We therefore report **availability of the ITS-format file at the registered URL** and do not describe a row without it as a breach.

---

## 4. Data

- **Register:** `OTHER.csv` (Title II, 1,026 rows after dropping two copies of the header line found inside the file) and `EMTWP.csv` (e-money tokens, 50 rows; 2 placeholders with no white paper are excluded); `ARTZZ.csv` (asset-referenced tokens) has no rows. The date field mixes `dd/mm/yyyy` (1,057 rows), `dd.mm.yyyy` (13 Latvian rows) and empty values (6).
- **Cohort:** rows whose "Last update of the record" ( `wp_lastupdate` ) is on or after 23 December 2025: **440 rows** (414 Title II, 26 e-money token), 425 distinct URLs. One row is dated 02/12/2026, after the register's own update; it is kept and flagged.
- **Cross-check:** the Wayback Machine capture of `OTHER.csv` of 2025-12-23 10:13 UTC has 625 rows and its latest date is 2025-12-22, consistent with the rule. Of the 414 Title II cohort rows, 403 have a key (LEI, CASP LEI, URL) absent from that capture and 11 are earlier entries with a new date. Of the current rows absent from the capture but outside the cohort, 58 carry an earlier date (mostly URL edits of older entries) and 3 carry none. There is no December capture of `EMTWP.csv`.
- **Taxonomy** `mica_taxonomy_2025.zip` (sha256 `19921398...aae033` ); **GLEIF** API (CC0), 238 LEIs.

---

## 5. Method

Full rules in `METHOD.md`, frozen before collection, with deviations D1–D11 appended. In short:

1. Every registered URL was fetched once on 2026-10-03 (cohort from 11:14:46 UTC, whole register by 11:49 UTC) with an identifying User-Agent, at most one request per second per host, no log-ins and no circumvention of access controls; on document hosts `robots.txt` is honoured as RFC 9309 specifies (D8). Anti-bot walls were recorded.
2. HTTP redirects and instant client-side redirects (D9) are followed; from an HTML page, at most three XHTML-like links and one PDF are fetched, one hop only.
3. Each document is classified from its bytes (Inline XBRL 1.1 with an ESMA MiCA entry point, Inline XBRL 1.0, other Inline XBRL, XHTML without Inline XBRL, PDF, ZIP, HTML, not retrievable); a row takes the best class reached.
4. `ixbrl-esma` files were validated with Arelle 2.46.0, offline, with the pinned package and formulas on; **valid** means no message at level ERROR or above, unsatisfied ESMA assertions of error severity included.
5. LEI and DTI facts were compared with the register row and LEIs looked up in GLEIF. Documents were deleted after processing.

**Deviations that move the figures** (details in `METHOD.md`): D1 reads a 401/403 on `robots.txt` as "unavailable" (RFC 9309), not as "not retrievable"; D7 parses all date formats of the register; D8 applies `robots.txt` with RFC 9309 wildcards; D9 follows instant client-side redirects; D10 separates Inline XBRL 1.0. The figure under the frozen access and redirect rules (no D1, no D9; D7, D8 and D10 correct errors and apply to both) is given beside the headline.

---

## 6. Results

### 6.1 Availability

| OUTCOME AT THE REGISTERED URL ( $\leq 1$ HOP)        | COHORT ROWS | REST OF REGISTER |
|------------------------------------------------------|-------------|------------------|
| Inline XBRL 1.1, ESMA MiCA taxonomy                  | <b>143</b>  | 20               |
| HTML page, no document link found                    | 129         | 117              |
| PDF only                                             | 69          | 411              |
| HTTP error (4xx/5xx)                                 | 30          | 43               |
| anti-bot wall                                        | 24          | 6                |
| XHTML without Inline XBRL                            | 20          | 3                |
| network error or <code>robots.txt</code> unreachable | 15          | 23               |

| OUTCOME AT THE REGISTERED URL ( $\leq 1$ HOP)     | COHORT ROWS | REST OF REGISTER |
|---------------------------------------------------|-------------|------------------|
| disallowed or reserved by <code>robots.txt</code> | 4           | 6                |
| Inline XBRL, other or wrong taxonomy reference    | 3           | 3                |
| ZIP archive                                       | 2           | 0                |
| Inline XBRL 1.0 (not the required version)        | 1           | 0                |
| register field is not a URL                       | 0           | 2                |
| <b>total</b>                                      | <b>440</b>  | <b>634</b>       |

**143 of 440 register rows (32.5%; 95% CI 28.3–37.0%)** lead to an ITS-format file, which correspond to 141 distinct files: 51 rows at the registered URL itself and 92 one link away from a landing page (2 of them after an instant client-side redirect). Under the frozen access and redirect rules the figure is **139 of 440 (31.6%; 95% CI 27.4–36.1%)**; D1 changed 45 rows (2 of them now reach the format) and D9 changed 2. Restricted to the 403 Title II rows that are new since the December capture, the figure is 141 of 403. Of the 26 e-money-token rows, 1 of 26 reaches the format; 16 are PDFs. The intervals are Wilson score intervals.

73 of 440 rows (16.6%) ended in something we could not read. The three "wrong taxonomy" files reference non-existent ESMA-like paths ( `.../taxonomy/2025/mica/mica-all.xsd` , `.../taxonomy/2025-03-31/mica/mica.xsd` ) or a self-hosted copy of the taxonomy; the Inline XBRL 1.0 file (OTHER line 519) uses the 2008 namespace with ESMA's entry point. Four available files could not be re-read for the version check (D10) and keep their first class.

There is no steady trend over the period: by quarter of the register update, 48 of 177 rows reach the format in 2026-Q1, 50 of 121 in Q2 and 45 of 137 in Q3 (0 of 4 in the last days of 2025 and 0 of 1 dated in 2026-Q4). By home Member State: Germany 77 of 101, Ireland 41 of 126, Malta 17 of 65, Luxembourg 2 of 47, the Netherlands 1 of 42, France 0 of 15. The German figure is entirely the main producer's (§6.3).

**One crypto-asset service provider.** On 2026-10-03, for the 38 cohort rows that name Bitstamp Europe S.A. as crypto-asset service provider, we observed an Incapsula anti-bot page on 20 rows ( `www.bitstamp.net` ) and, on 18 rows, XHTML files on `assets.bitstamp.net` titled "MiCA Whitepaper Inline XBRL" that declare no Inline XBRL namespace and contain no XBRL fact. The same title appears on 19 Inline XBRL files in ESMA's taxonomy served by eight other hosts (seven on `socios.com` ), so it is a label of a common template or authoring tool. Under the frozen rule these 18 rows were "robots" ( `assets.bitstamp.net` answers 403 to `/robots.txt` ) and are reported thanks to D1. We describe only what the registered URLs served; we cannot tell what was notified to the authority.

## 6.2 Validity

**110 of 143 rows** point to a file that validates with zero errors, i.e. **108 of 141 distinct files**. All files are Table 2 (other crypto-assets) except two Table 4 (e-money token) files; Arelle evaluates 216 ESMA assertions on each Table 2 file and 109 on each Table 4 file. Of the 33 rows with invalid files, 16 fail only ESMA assertions and 17 have XBRL or Inline XBRL errors (schema value errors, Inline XBRL transformation

errors), usually together with assertion failures. The assertions most often failed at error severity are the unit/token rules for fields E.3–E.5 (8–9 rows each) and the country rule `A6ConditionalRuleA4Country` (6); some may be stricter than the ITS itself. Across the 153 distinct files of available rows in the whole register, 113 of 153 validate and 151 of 153 are well-formed XML with an XHTML root, i.e. a single XHTML file in the sense of Art. 2(1)(a).

Files reached directly at the registered URL validate less often (24 of 51 rows) than files one link away (86 of 92), because most one-hop files come from the main producer.

### 6.3 Who produces the valid files

The host that served each file was grouped by registrable domain. **77 of 110 valid rows** (70.0%; 76 distinct files) are served by `crypto-risk-metrics.com`, the white-paper site of Crypto Risk Metrics GmbH. It is itself the offeror or person seeking admission named in the register on those rows: the register LEI equals the offeror LEI tagged in the file in 77 of 77. All 77 of its available rows validate. The register names it on 86 cohort rows; the other 9 end on landing pages without a link to a document. The other 66 available rows come from 52 other domains, and 33 of 66 validate; 8 of those rows are on generic hosting platforms (`storage.googleapis.com`, `github.io`, `r2.dev`), where the domain names the platform, not the producer. Outside the one specialised producer, the format is rarer and valid in half the cases.

### 6.4 Identifiers

For 2 available rows the identifiers could not be read (D6). Of the rest:

- **LEI.** The register gives no LEI for 39 of the available rows. Where it does, the register LEI appears among the file's LEI facts or context identifiers in **103 of 103** rows. All these register LEIs exist in GLEIF; 2 are lapsed. In the files, 26 rows tag no offeror LEI (the ITS requires it "where available") and 3 tag an offeror LEI that is lapsed in GLEIF.
- **DTI.** A register DTI appears in the file in **134 of 136** rows; the register's DTI-FFG in 130 of 134.
- **Dates.** 6 files tag a date of notification before 23 December 2025 although the register dates the rows in 2026: the cohort rule includes some white papers notified shortly before the ITS applied, which some preparers already produced in the new format.

### 6.5 Context

The rest of the register (634 rows dated before 23 December 2025 or undated) reaches the format in **20 of 634 rows**; 411 are PDFs, as expected before the ITS applied.

---

## 7. Limitations

- **Availability, not compliance.** We did not see what national authorities received; a missing file at the URL does not prove a missing file at the authority (§3).
- **One day.** Results describe what each URL returned on 2026-10-03 to a polite, identified client; files change (12 files out of 168 had changed when downloaded again the same afternoon), and `robots.txt` was re-read in the afternoon.

- **One hop, static HTML.** Pages that build their links with JavaScript, or link the file two clicks away, count as "HTML page, no document link" (e.g. 25 rows on one exchange's "learn" pages). The 129 such rows are an upper bound on "no file at all". Instant client-side redirects are followed (D9).
  - **Cohort definition.** `wp_lastupdate` is a register field ("Last update of the record"), not the notification date; a register correction can bring an older white paper into the cohort, and register lag can bring in white papers notified in late 2025 (6 seen among the files).
  - **Walls and exclusions** (28 rows) were not circumvented and may hide files in the format. **Validation** is as good as Arelle and ESMA's assertions; Annex I content quality and warnings are not assessed.
  - **Interval.** The Wilson interval treats the cohort as a sample of a process; as a census on one date, the count is exact.
- 

## 8. A checker anyone can run

`scripts/mica_wp_check.py <file-or-url>` prints a JSON verdict on one white paper (class, Inline XBRL version, single XHTML, entry point, Arelle with ESMA's assertions, LEI checksum, optional LEI/DTI match), each finding citing the ITS or the taxonomy, in about 5 seconds offline. Preparers hold the file before notifying it, so the input exists. EasyxLab may release it as a free tool ( `mica-wp-lint` ).

---

## 9. Data, code and licences

`data/rows.csv` (one line per register row, with the registered `wp_url` , the URL of the document classified, the outcome under current and frozen rules, version, validity and identifier matches), `data/documents.csv` , `data/entities.csv` and `data/metrics.json` hold every figure; `scripts/run.sh` recomputes them offline and `scripts/check_headline.py` re-derives the cohort and classes and asserts them. Code Apache-2.0; data and text CC BY 4.0; register-derived data reused with acknowledgement of ESMA; GLEIF data CC0. No white paper is redistributed, and no personal data was extracted.

---

## 10. Automation and review

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

- **Done by an AI agent:** the prior-work search; downloading and quoting the legal texts, ESMA's 51 MiCA Q&As, its statement and manual; defining and freezing the cohort and classification rules; writing the fetcher, the collector, the checker and its tests; the collection, validation, GLEIF look-ups and analysis; the deviations; writing this paper, `METHOD.md` , `README.md` and `VERIFICATION.md` .
- **Checked by an independent AI reviewer:** a second agent, asked to be adversarial, recomputed the figures, verified the legal quotations, re-fetched 50 cohort rows (47 agreed) and re-validated 16 files with Arelle (the same verdict for all 16). It found four blocking problems — date formats and header lines in the register (D7), Inline XBRL 1.0 counted as the required format (D10), a `robots.txt` parser without RFC 9309 wildcards (D8) and instant redirects counted as pages without a document (D9) — and several reporting issues; all were corrected and the figures above are the corrected ones.

No issuer, service provider, authority or data provider was contacted, and nothing was submitted anywhere.

None. EasyByte, the cooperative behind EasyxLab, has no product or service related to MiCA. EasyxLab may release the checker written for this study as a free tool ( `mica-wp-lint` ).

## References

1. Regulation (EU) 2023/1114 (MiCA), consolidated text CELEX 02023R1114-20240109; Regulation (EU) 2023/2869.
2. Commission Implementing Regulation (EU) 2024/2984, CELEX 32024R2984.
3. ESMA, MiCA Q&As (Q&A tool, read 2026-10-03), incl. Q&A 2845 and 2654.
4. ESMA, Statement ESMA75-1303207761-6284, 28 November 2025; MiCA XBRL taxonomy 2025 and reporting manual v1.0.
5. ESMA, interim MiCA register (OTHER.csv, EMTWP.csv, ARTZZ.csv), last update 30 September 2026.
6. RFC 9309, Robots Exclusion Protocol, 2022. Arelle, `arelle-release` 2.46.0. GLEIF LEI records API.
7. sebastianfoerste/micar-register-observatory; sebastianfoerste/micar-whitepaper-linter; Loideroi/WhitePaperXBRL (GitHub, READMEs read 2026-10-03).
8. Asset-Class Specific Sustainability Disclosure: Lessons Learned from the EU MiCA Regulation, arXiv 2609.26932, 2026.
9. Legal NLP Meets MiCAR: Advancing the Analysis of Crypto White Papers, NLLP 2023, doi:10.18653/v1/2023.nllp-1.14.

## Cite this study

### CITATION

EasyxLab (2026). Are MiCA crypto-asset white papers published in the machine-readable format required since 23 December 2025? Study S15. EasyByte Hub S. Coop. Mad. <https://github.com/easybytehub/easyxlab/tree/main/studies/s15-mica-white-papers>

## BIBTEX

```
@techreport{easyxlab_s15,  
  title      = {Are MiCA crypto-asset white papers published in the machine-  
readable format required since 23 December 2025?},  
  author     = {{EasyxLab}},  
  institution = {EasyByte Hub S. Coop. Mad.},  
  number     = {S15},  
  year       = {2026},  
  url        = {https://github.com/easybytehub/easyxlab/tree/main/studies/s15-  
mica-white-papers}  
}
```