

Who publishes the numbers

When a Greek business looks up a number about its own market, where does that number come from, and now that a machine chooses which one to repeat, whose gets repeated?

Nikolaos Broikos · broikos.gr First edition · v1.0

68 programmes · 64 publishers · 15 sectors

68

PUBLICATION PROGRAMMES CENSUSED

26.5%

BUILT ON DATA THE PUBLISHER ALREADY OWNS

2.9%

SIT BEHIND AN EMAIL FORM

10%

FRAMED AS CSR OR CONTRIBUTION

I run businesses in Greece, and almost every decision I make rests on a number somebody else published. What the market is worth. How fast it is growing. What share of it moved online. I have quoted those numbers in pitches, priced against them, and planned a year around them, without once being able to say who produced them, why, or whether anybody had ever checked.

So I set out to answer a question that turns out to have no published answer: **when a Greek business looks up a number about its own market, where does that number actually come from?**

The short answer is that it comes from a company with something to sell. In Greek short-term rentals the definitive figure is published by a property-management software company. In residential property prices it is a listings platform. In ship finance it is a small consultancy that has kept an index running since 2001. **None of them is a statistical agency, all of them have a commercial interest in the subject, and all of them give the number away for free.**

That is not a scandal. It is an arrangement, and it is the one the Greek economy actually runs on. But an arrangement nobody has described cannot be judged, and I wanted to judge it before trusting it any further.

Why this is a 2026 question and not a 2019 one

Until recently, a published number reached a decision-maker through a journalist. Somebody chose it, attributed it, and put it in a sentence a reader could trace. That path is now the minority path.

A Greek business owner asking what the e-commerce market is worth increasingly asks a model, and **the model returns one number with a citation attached**. The editorial layer has been replaced by a retrieval layer, and the retrieval layer has different tastes: it does not care about a press relationship, it cannot be pitched, and it will happily repeat a vendor's marketing figure with the confidence of a national statistic.

Which means the question “who publishes the numbers” has quietly become **“whose numbers does the machine repeat”**, and I could not find anyone who had measured that for Greece. So this paper does both: it censuses who publishes, and then it asks three answer engines thirty-six questions about the Greek economy and codes every source they cite.

What this paper is the foundation for

This is the first of seven, and it exists to establish the standard the other six are held to. If the finding is that Greek business research is dominated by commissioned surveys with undisclosed methods, then a paper claiming to do better has an obligation to show its workings, publish its data, and record what failed verification. **Everything the other six papers do, the claim ledgers, the published instruments, the retractions in the body, is a response to what this census found.**

It is also, deliberately, a test of whether the category is open. It is: only a quarter of Greek publication programmes are built on data the publisher already owns, against a considerably higher share internationally. That gap is the reason the other six papers were worth writing at all.

Somebody publishes the number your industry quotes. In Greek short-term rentals it is a property-management software company. In residential property prices it is a listings platform. In ship finance it is a small consultancy that has kept an index running since 2001. None of them is a statistical agency, and all of them give the number away.

This is the first census of that behaviour in Greece that I am aware of: **68 publication programmes from 64 publishers across 15 sectors**, assembled through 36 documented searches, coded on fourteen variables^{1,2}, and checked for coding reliability three times, once by an independent coder.

FINDING 01

Greek publishers under-use the data they already own

Only 18 of 68 programmes are built on the publisher's own operational data. Against a comparable international corpus on a matched schema, Greece runs 31.6% against 42.7%⁸: an eleven-point deficit. Commissioned surveys are equally dominant in both.

FINDING 02

Almost nobody gates

Two programmes out of 68 sit behind an email form, and both belong to software companies selling outside Greece. Internationally the figure is 13.4%, more than four times higher.

FINDING 03

It is not CSR

Seven of 68 programmes are framed as contribution, footprint or responsibility: the one coding column that survived independent replication at 89.7% agreement. Category ownership, at 23 programmes, is the most common motive coded, though motive itself proved the corpus's least reproducible column.

FINDING 04

Fewer than half show their workings

Thirty of 68 disclose method fully. Six disclose nothing at all: asking readers to take a number entirely on trust.

FINDING 05

The press picks up numbers, not documents

Across 25 programmes tracked, deep practical whitepapers drew zero Greek media coverage. Own-data releases averaged 4.6 outlets; free summaries over a paid dataset averaged 6.0.

FINDING 06

English costs you Greece

Programmes published only in English averaged 0.8 Greek outlets. Greek or bilingual programmes averaged 4.4 to 4.5: more than five times as many.

This is a paper about corporate research, published by somebody who produces corporate research, and a reader is entitled to hold that against it. My defence is in section 3, where I print the coding failure that nearly put a wrong number on the cover.

03 **About this research**

I censused organisations that are Greek or substantively Greek-operating, that publish **original** research, their own data, their own survey, or a study they commissioned, and that make it free in whole or as a substantive teaser. Translated global editions, news roundups and paid-only research were excluded.

The corpus was built by systematic search across fifteen sectors in Greek and English, by walking the member lists of six business associations, by sweeping Greek business-press archives for “έρευνα της [εταιρείας]” patterns, and by snowballing from co-publishers and named institutes. Every query is logged with its date and its yield, **including the eleven that returned nothing**.

State bodies are excluded throughout. ELSTAT, the Bank of Greece, EETT and the regulators publish more free research than any company here, but statutory publication is not the behaviour being studied. Business associations funded by their members are included, because that is the same behaviour with the cost pooled.

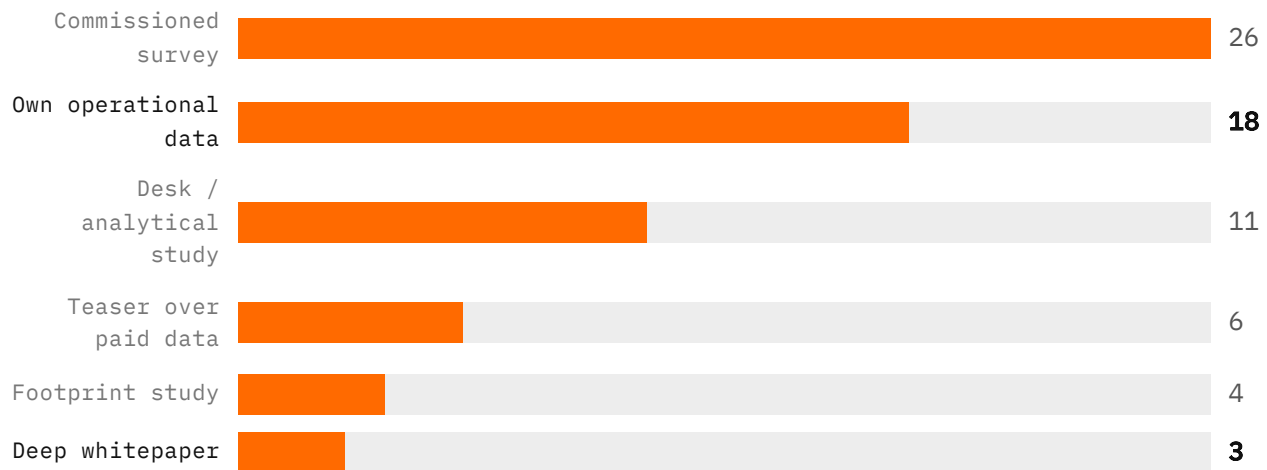
The unit is the **publication programme**, a publisher plus a recurring title, not the organisation and not the individual edition. One consultancy runs three programmes with different formats and audiences; collapsing them to one row would lose the variation this paper is about.

Motive is **inferred from four observable signals**, never asserted: who authored the artefact, whether it is gated, what its headline metric measures, and who it addresses. The decision rule is published so any row can be re-derived or disputed.

Figures are descriptive. No claim of statistical significance is made anywhere: the corpus is what systematic search surfaces, not a random sample of Greek companies. Every count is a floor and is written as “at least”.

04 **What they publish**

Six formats appear in the corpus.



Only about one programme in four is built on data the publisher already owns. Fifty-four per cent are surveys of respondents or analysis of somebody else's figures.

To test whether that is a Greek trait or simply what corporate research looks like everywhere, I coded the same variables across a comparison corpus of 82 European and American publishers, matching the schemas by setting the Greek desk-analysis rows aside.

FORMAT MIX, GREECE AGAINST AN INTERNATIONAL COMPARISON CORPUS

FORMAT	GREECE (N=57)	INTERNATIONAL (N=82)
Own operational data	31.6%	42.7%
Commissioned survey	45.6%	43.9%
Teaser over a paid dataset	10.5%	4.9%
Institute-authored footprint study	7.0%	1.2%
Deep practical whitepaper	5.3%	7.3%
Email-gated (all rows)	2.9%	13.4%

The result is more specific than the intuition I started with. **Commissioned surveys are equally dominant in both corpora**, so buying a survey is not a Greek habit; it is the global default. What is distinctively Greek is the eleven-point deficit in own-data research, the four-and-a-half-fold gap in gating, and an unusually heavy reliance on institute-authored footprint studies: almost six times more common here.

Greek companies are sitting on datasets and commissioning surveys about the same subjects.

The comparison corpus is a convenience sample assembled by a looser protocol than this census, coded by the same person on an earlier version of the schema. **It supports a direction, not a precise gap.**

Fewer than half show their workings

“Fully disclosed” is not a demanding bar. It requires stating what the data is, what period it covers, and how it was collected. Thirty of 68 clear it. Thirty-two disclose partially. Six disclose nothing at all.

A FINDING ABOUT THIS PAPER

While coding this corpus I ran a reliability check, re-coding a fifth of it blind. **It failed, at 0.64 against a 0.80 threshold.**

The reason was a missing code. Bank and institute *desk analyses*, original work on third-party data with no new collection, had nowhere to go in the schema and had been defaulting into the commissioned-survey category. Eleven programmes sat in the wrong bucket, and the resulting headline, “36 of 70 programmes are commissioned surveys”, was wrong by ten.

I added a sixth format, recoded the corpus, drew a fresh sample and re-checked, reaching 0.93. The corrected figure is 26 of 68.

I am printing this because a paper whose subject is the transparency of other people's research has no standing to conceal a defect in its own, and because it carries an obvious implication for the six programmes that disclose no method at all.

Nobody catches what nobody can see.

05 Why they do it, and the CSR question

PRIMARY MOTIVE, INFERRED FROM FOUR OBSERVABLE SIGNALS

MOTIVE	PROGRAMMES	SHARE
Category ownership	23	33.8%
Policy	13	19.1%
Demand generation	13	19.1%
Teaser to paid	6	8.8%
PR amplification	5	7.4%
CSR marketing	4	5.9%
Licence to operate	3	4.4%
Employer brand	1	1.5%

The single most common reason a Greek company publishes research is to own the recurring number in its category. A third of the corpus. Not leads, not reputation in the abstract, not social contribution: becoming the organisation that gets called when a journalist needs the figure.

So: is it CSR?

No. Seven programmes of 68, 10.3%, are framed as contribution, footprint or responsibility. Licence to operate is the *primary* motive for three programmes, 4.4% of the corpus.

And those three have an identical shape. Each is an **institute-authored footprint study funded by a manufacturer or an industry association**. The company pays, an independent institute writes and signs, the output is measured in jobs and multipliers and share of GDP, nothing is gated, and the audience is the state.

Read plainly, these are arguments addressed to government: this sector is too economically load-bearing to tax, restrict or ignore. That is legitimacy-seeking, and it is a perfectly respectable thing for an industry to do. It is not philanthropy, and treating it as philanthropy misreads both the document and the deadline it was written against.

Research motivated by licence to operate defends the publisher's right to operate. Research motivated commercially defends its right to be considered.

Sixty-one of the 68 programmes are the second kind.

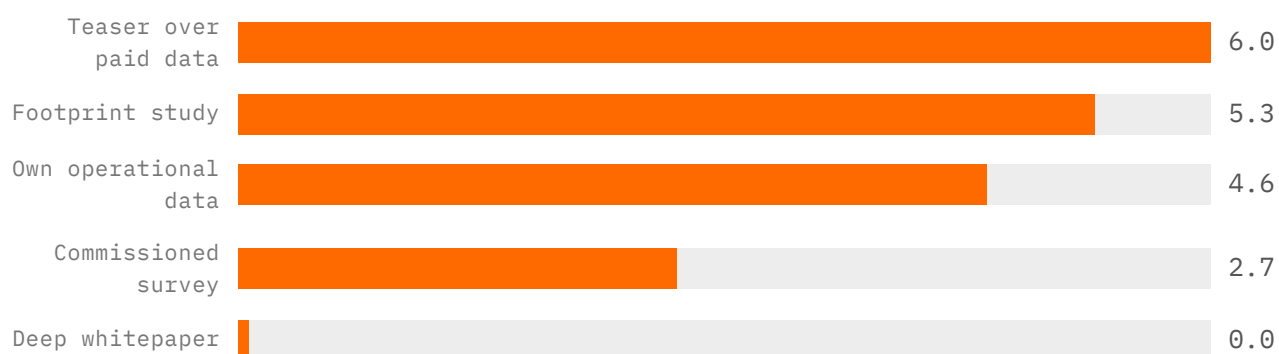
There is a second, quieter pattern. Four programmes are coded CSR-marketing: research genuinely about a social or environmental question that also happens to promote the publisher's own programme in that area. A beverage company surveying whether consumers will pay more for sustainable hospitality, having a sustainable-hospitality scheme to sell. A telecoms operator measuring the digital maturity of the small businesses it sells digital services to. These are not cynical, and they are not disinterested either. The honest description is that the question is real and the choice of question is commercial.

06 Does it work in Greece?

I measured one thing: how far a publication travels in the Greek press. For 25 programmes stratified across the formats, I ran a single dated search and counted the distinct Greek outlets carrying the latest edition.⁴

This is a **floor**, and a rough one. It counts online press only, misses print and broadcast entirely, and in seven of 25 cases it hit the search-result ceiling, so the strongest performers are undercounted. It separates low pickup from high pickup, not high from higher.

Overall: **median 4 outlets, mean 3.7.**



Deep whitepapers drew no Greek press coverage at all. Both cases in the sample, zero. That includes a well-made 47-page document built on more than a million records, whose publisher's *statistics release* in the same year drew four outlets. Same company, same data asset, same year: the number travelled and the document did not.

The Greek press covers findings, not artefacts. A whitepaper is bought with attention from the reader who already has the problem; it is not a press play, and expecting it to be one is how a good document gets mistaken for a failure.

The most-used format is among the weakest performers. Commissioned surveys are 38% of the corpus and average 2.7 outlets. The ranges overlap and the samples are small, so this is a tendency rather than a rule, but the tendency runs against the corpus's own investment pattern.

English costs you Greece

MEDIA PICKUP BY PUBLICATION LANGUAGE

LANGUAGE	PROGRAMMES	MEAN GREEK OUTLETS
Greek and English	10	4.5
Greek only	10	4.4
English only	5	0.8

Roughly a five-fold difference. Publishing bilingually costs a translation and appears to lose nothing; publishing only in English forfeits the Greek press almost entirely. For a company whose customers are Greek, this is the single most actionable line in the paper.

Giving it away outperforms gating it

Programmes coded demand-generation averaged **zero** Greek outlets across three cases. Programmes coded teaser-to-paid averaged **6.0** across three.

Both models withhold something. The difference is what. The teaser withholds the *dataset* and publishes the *findings*, so there is something to report. The gate withholds the findings, so there is not. In a market this size, where the same fifteen or twenty business outlets carry everything, that appears to be the difference between existing and not.

Repetition compounds, modestly

PICKUP AGAINST PROGRAMME AGE

EDITIONS TO DATE	PROGRAMMES	MEAN OUTLETS
1-6	9	3.0
7-20	6	3.8
21+	10	4.3

Monotonic and in the expected direction. The gradient is gentle and edition counts are mostly estimated, so this is a pattern rather than a coefficient. But it matches what the long-running publishers look like from outside: they are not covered because any single edition is remarkable, they are covered because they are the series of record.

And the machines that now answer the question

Press pickup measures whether a journalist repeats your number. That is a proxy for something which now matters more and can be measured directly: whether a *machine* repeats it. The twelve prompts for this test were fixed and published on 5 August 2026, before any engine was touched, so the protocol could not be tuned to a flattering result.

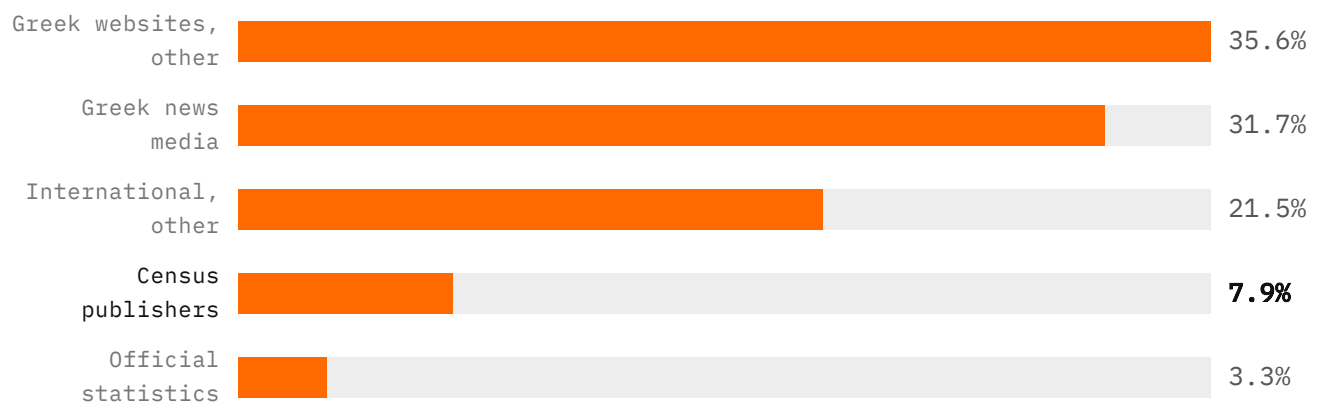
On **6 August 2026** I put those twelve questions, in Greek, as a Greek user would type them, to three answer engines with web access, one session each, first response only, no follow-ups. **Thirty-six answers naming 303 sources.**

M2: ANSWER-ENGINE CITATION TEST, 6 AUGUST 2026. ONE RUN, TWELVE FIXED PROMPTS PER ENGINE.

ENGINE	SOURCES PER ANSWER	NAMED A CENSUS PUBLISHER	NAMED ANY GREEK SOURCE
Perplexity sonar	13.8	8 of 12	12 of 12
ChatGPT gpt-5-search-api	4.5	6 of 12	11 of 12
Gemini 2.5-flash, Google Search	6.9	5 of 12	11 of 12
All three	8.4	19 of 36: 52.8%	34 of 36: 94.4%

Across all 36 answers, **52.8% named at least one publisher from the census, and 44.4% named the specific publisher that owns that sector's recurring number.** Almost every answer, 94.4%, named a Greek source, so this is not engines falling back on international material for want of anything local.

The share of *citations* is the sobering figure. Of the 303 sources named, **only 24, 7.9%, were census publishers.**



Source mix across all 303 citations, classified by domain against the census. Full log in ai-citations.csv.

The engines mostly cite the press *reporting* the research rather than the research itself. A corporate publisher's number reaches the answer through a news site that quoted it, which means the publisher gets the number into circulation and the news site gets the citation.

Where a company owns the recurring number, every engine named it. Where nobody owns it, no engine named anyone.

The sector pattern is the actual finding. The prompts were chosen in advance to cover sectors where the census found an obvious owner of the number, plus one control sector where it found none.

WHETHER THE SECTOR'S CENSUS OWNER WAS NAMED, BY PROMPT. THREE ENGINES, SO THREE CHANCES PER ROW.

QUESTION	SECTOR	OWNER NAMED
How many vehicles were registered in Greece this year?	Automotive	3 of 3
What does brewing contribute to the Greek economy?	Food and drink	3 of 3
How big is the Greek pharmaceutical market?	Pharma	3 of 3
What is average occupancy in Greek short-term rentals?	Short-term rental	2 of 3
How much did Greek house prices rise last quarter?	Real estate	1 of 3
How is Greek shipping financed by the banks?	Shipping	1 of 3
What does a software engineer earn at a Greek startup?	Startups	1 of 3
Who are the most attractive employers in Greece?	HR	1 of 3
How big is the Greek e-commerce market?	E-commerce	1 of 3
What share of Greeks buy online, and how often?	E-commerce	0 of 3
How many ransomware attacks hit Greek businesses?	Cybersecurity	0 of 3
What does industrial electricity cost in Greece versus Europe?	Energy: control, no owner	0 of 3

Three sectors returned their owner in every answer from every engine. The control returned no corporate publisher at all, from any engine, which is what the design predicted and is the cleanest evidence in this paper that ownership of a number is a real and machine-visible property, not a story publishers tell about themselves.

E-commerce is again the largest miss. It is the sector with the most Greek publishers and the loudest trade press, and across six answers about market size and shopper behaviour its nominal owner surfaced once.

One organisation dominates the citations. IOBE, the Foundation for Economic and Industrial Research, was named in **11 of 36 answers**, through three separate commissioned programmes: brewing, food and drink, and pharmaceuticals. No other publisher appears more than three times. The most citable object a Greek company can pay for is an independent institute's name on a study of its own sector's economic contribution. That is the same format that led press pickup at 5.3 outlets, and it is the format this corpus produces least.

What this test is and is not. One date, one run, three engines, twelve prompts. Answer engines are non-deterministic and personalised; a re-run would give different specifics. This is a snapshot, not a benchmark, and the claim it supports is narrow: that corporate research surfaces in generated answers *at all*, and surfaces far more reliably where a publisher owns the recurring number.

The locked protocol specified four engines. **Microsoft Copilot was not run**, no API access, and that is recorded rather than quietly redefining the design as three. One Gemini response returned an API error and is counted as an answer naming no sources, not dropped. Sources were matched to the census mechanically by domain, so a publisher named in prose but not linked is not counted; the 52.8% is therefore a floor. Full log, prompts and every raw response are published.

A separate observation, outside this protocol. While using the same engines as a finding aid for an unrelated source, one returned a verbatim Greek-language “methodology note” in quotation marks and attributed it to a specific PDF. I downloaded the PDF. The quoted sentence is not in it. That cost nothing here because nothing entered this paper without being opened at source, but it is the reason nothing does.

An earlier draft carried two external statistics about why published research works commercially. Both were cut at citation audit: one could not be opened at source, and the other, the widely-quoted claim that only 5% of buyers are in-market at any moment, turned out to originate not with the report I had attributed it to but with separate research quoted inside it. The claim may well be sound. My attribution was not, and a paper that spends a section on other people's method disclosure does not get to publish a citation it has not opened.

07 **How Greece differs from everyone else**

The Greek census has a companion: a lighter scan of 119 programmes across Greece, Europe and America, coded on a matched schema. It is not the same quality of instrument and is used here only for shape. On that basis three differences stand out, and one of them explains most of the others.

What this corpus is, and what it is not. The 119-row comparison set was assembled by systematic search across three regions with a deliberately lighter verification standard than the Greek census: **64 of the 119 rows are confirmed and 55 are unverified.** It carries no reliability check and no second coder.

It is therefore used for one purpose only: comparing the *distribution* of formats and access models between regions, where a consistent error would have to be regionally biased to matter. No count from it appears as a headline in this paper, and no Greek claim rests on it. Where the Greek census and this corpus disagree, the census wins.

America publishes its own data. Greece commissions surveys.

FORMAT DISTRIBUTION ACROSS THE 119-PROGRAMME COMPARISON CORPUS. PERCENTAGES ARE OF EACH REGION'S OWN TOTAL: GREECE 37, EUROPE 36, AMERICA 46.⁸

FORMAT	GREECE	EUROPE	AMERICA
Platform benchmark: the publisher's own operational data	19%	19%	48%
Commissioned survey	43%	47%	30%
Research programme	8%	3%	9%
Benchmark, other	8%	11%	4%
Deep white paper	5%	17%	0%
Free teaser over paid data	8%	3%	7%
Footprint study and advocacy	8%	0%	2%

Nearly half of the American programmes are built on the publisher's own platform data, against a fifth in Greece and a fifth in Europe. The mirror image is the commissioned survey: 43% of Greek programmes and 47% of European ones, against 30% American.

This is the same finding the Greek census produced from a better instrument, 18 of 68 programmes on own operational data, arriving from a different direction. Greek publishers buy their evidence. American publishers extract theirs from the systems they already run.

The economics of that difference are not subtle. A commissioned survey is a recurring external cost with a fieldwork lead time, and it produces a number anybody else can also commission. Own operational data has no fieldwork cost, no lead time, and cannot be replicated by a competitor at any price. The Greek corpus is concentrated in the expensive, replicable format and thin in the cheap, defensible one.

**The format Greek publishers use most is the one they cannot own.
The format they use least is the one nobody can take from them.**

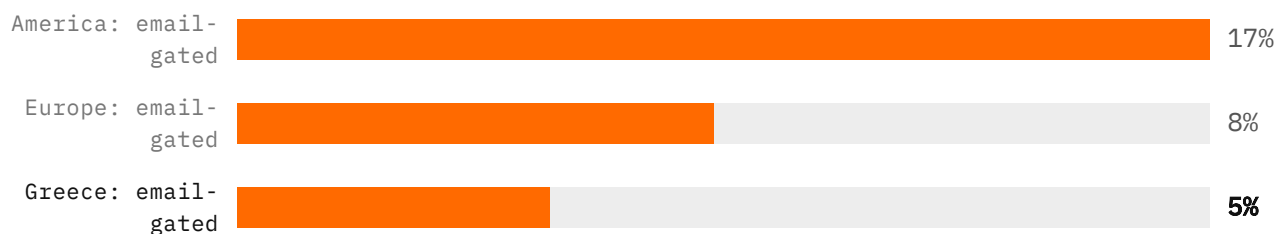
The deep white paper is a European habit

The document type that gives this genre its name barely exists outside Europe in this sample: **17% of European programmes, 5% of Greek, and none at all of the 46 American ones.**

Set that beside the Greek media result reported earlier, deep practical whitepapers drew zero Greek press coverage across both cases in the amplification sample, and a consistent picture appears. The long teaching document is a real format with a real audience, and that audience is a reader with a problem rather than a journalist with a deadline. American corporate publishing, which is the most commercially optimised of the three, has largely stopped producing it.

That is not an argument against writing one. It is an argument against expecting it to travel. A whitepaper is bought with the attention of somebody already looking for the answer; it is not a distribution strategy, and the American corpus suggests the market worked that out some time ago.

Gating is an American practice



Share of each region's programmes placing the full document behind an email form. Comparison corpus, n=119.

American publishers gate more than three times as often as Greek ones. The Greek census puts the domestic figure lower still, at two programmes in 68, and both of those belong to software companies selling outside Greece.

Two readings are available and the data cannot separate them. Either Greek publishers have understood that gating suppresses the reach that makes the exercise worthwhile, or they have not built the lead-capture machinery that makes gating worth doing. The Greek amplification result, ungated programmes reaching more Greek outlets than gated ones, is consistent with the first, and the near-total absence of marketing-technology publishers from the Greek corpus is consistent with the second.

The sectors that do not exist in Greece

The comparison corpus carries 31 sector labels that appear in Europe or America and not once in Greece. They are not random. They cluster into three groups, and the clustering is the point.

- 01 **Infrastructure and network operators:** internet infrastructure, telecom security, network traffic. Companies whose product generates a global dataset as a by-product, and who publish it because it is free to produce and impossible to contest.
- 02 **Payments and insurance:** payments/retail, payments/e-commerce, insurance, reinsurance. Sectors that sit on transaction and loss data and, abroad, publish from it routinely.
- 03 **Developer and marketing technology:** developer tools, marketing tech, B2B marketing, media/advertising. The sectors that invented gated research, absent from the Greek corpus almost entirely.

The Greek census found publishers in Greek payments and Greek banking, so the second group is not empty in reality: it is thin. The first and third are close to genuinely absent, and both are sectors where the publishable asset is a by-product of operations rather than a purchase.

Which returns the argument to where the census left it. The gap between Greek and international corporate research is not mainly a gap in effort, budget or sophistication. **It is a gap in whether the publisher looks at the data they already hold and recognises it as a publication.** Every sector in the first group abroad is a sector whose Greek equivalent runs the same systems and generates the same logs.

08 The white space

If a sector has no owner of its recurring number, the position is open.

Single-publisher sectors in this corpus: payments and energy. One programme each. Any credible second entrant with real data would be competing against a near-vacuum.

Sectors searched that returned nothing at all: industrial manufacturing footprint studies, courier and last-mile logistics, consumer electronics retail, gambling and gaming, and telecoms beyond a single university-partnered study. In courier and last-mile the absence is conspicuous, because the parcel operators sit on exactly the kind of operational data that produces a category-owning index, and the Greek e-commerce press covers the topic constantly using figures scraped from filed accounts and regulator reports.

If you are deciding whether to publish

- 01 **Check whether your category is taken.** If somebody has run the reference number for a decade, entering means a different angle or a better dataset: not the same survey with your logo on it.
- 02 **Prefer the data you own.** It is cheaper, faster, cannot be copied, and in this corpus it travels roughly seventy per cent further than a commissioned survey.
- 03 **Publish in Greek, or bilingually,** unless your customers genuinely are not here.
- 04 **Separate the news from the document.** Publish the finding openly and let the depth sit behind it. Treating a whitepaper as a press release produces a document nobody reports.
- 05 **Budget for the third edition.** One edition is a blog post. The organisations that own their categories here have been at it for twelve to twenty-four years.

One edition is a blog post. Three is an index.

09 Limitations

LIMITATION	WHAT I DID ABOUT IT
The corpus is what systematic search surfaces. Publishers with no web presence are invisible to it	Search protocol and all 36 queries published, including the 11 that found nothing ² . Every count written as “at least”
Motive is inferred from artefacts, not stated intent	Decision rule published; CSR framing coded separately from motive so the CSR answer can be re-derived without trusting the motive column
Media pickup counts online Greek press only, one dated search per programme, censored in 7 of 25 cases	Reported as a floor throughout, and used only to compare formats, never to score individual publishers
Several sub-findings rest on two or three cases. The zero-pickup result for whitepapers is n=2	Sample sizes printed in the sentence, not hidden in a note
Edition counts are mostly estimated from cadence rather than counted from archives	Flagged per row; no headline claim rests on an estimated count
The motive column failed inter-rater reliability at κ 0.34. Two coders working from the published rule agreed on why a programme was published just over half the time	Measured on all 68 rows and printed in full in Appendix B, including the two codes the rule had failed to define. Motive figures are labelled indicative. The corpus was deliberately <i>not</i> recoded afterwards. The CSR answer rests on csr_framed, agreement 89.7%, which was coded independently for exactly this reason
The independent second coder was a language model, not a trained human researcher	Stated. It tests whether the published rule reproduces a decision, which is what a published rule is for; it is not a substitute for a human second coder and the result is reported as the weaker instrument it is
The answer-engine test is one run on one date against three of the four engines specified	Prompts locked and published before the run; Copilot recorded as not run rather than the design being redefined; all 36 raw responses published
I publish research and would sell research services	Method, corpus and coding rules published in full; my own coding failure printed in section 3

The materially weakening limitation is the first one. This is a floor, not a total. Treat every count as the minimum true value, and the sectors reported as empty as sectors where I found nothing: not sectors where nothing exists.

10 Sector by sector: who owns the number

The census answers a question no Greek company can currently answer about its own market: in my sector, has somebody already become the source everyone quotes?

The table below lists every sector in the corpus, how many publication programmes it contains, and who, if anyone, publishes research built on data they already own. That last column is the one that matters, because own-data programmes are the ones competitors cannot replicate.

ALL 15 SECTORS, BY PROGRAMME COUNT

SECTOR	PROGRAMMES	PUBLISHERS	WHO PUBLISHES OWN-DATA RESEARCH
Tourism and hospitality	12	10	Athens International Airport, Hosthub, INSETE
E-commerce and retail	9	9	Skroutz, Wolt Greece
Banking, finance and insurance	6	6	EAE
Professional services	5	5	<i>nobody</i>
Real estate	5	5	Cushman and Wakefield Proprius, Danos / BNP Paribas Real Estate, Spitogatos
HR and labour	5	5	<i>nobody</i>
Cross-sector bodies	5	5	<i>nobody</i>
Food and beverage	4	4	<i>nobody</i>
Automotive and mobility	4	4	Ayvens Greece, Car.gr, SEAA
Startups and venture capital	4	3	Marathon Venture Capital
Shipping	3	3	Moore Greece, Petrofin Research
Pharma and health	2	2	<i>nobody</i>
Technology and cyber	2	2	Obrela Security Industries
Payments and fintech	1	1	Nexi Greece
Energy	1	1	<i>nobody</i>

Six sectors have nobody

22 of the 68 programmes sit in sectors where not one publisher builds on data they already hold. Professional services, HR and labour, the cross-sector business bodies, food and beverage, pharma and health, and energy: every one of those programmes is a survey, a desk analysis, a footprint study or a paid-data teaser.

That is not a criticism of those programmes. A commissioned survey is a legitimate instrument and some of the corpus's best work is survey-based. It is an observation about competitive position: in six sectors, the recurring number is up for the taking by whoever holds a dataset and decides to publish it.

Shipping is the smallest sector and the best run

Shipping contains three programmes. It also contains **the two longest evidenced runs in the entire corpus**, a ship-finance index running since 2001, and a ferry-market report in its 24th edition, and two of its three programmes are built on data the publisher compiles itself.

Both are small organisations. Neither is a statistical agency, a bank or a consultancy of any scale. They simply started early, published every year without exception, and became the reference. If the census contains a model worth copying, it is this one, and it did not require size.

E-commerce is the largest miss

E-commerce and retail is the second-densest sector at nine programmes. Only **two** are built on the publisher's own data.

This is the sector where every participant, marketplace, courier, payment processor, platform, merchant, holds transaction-level data by default. It generates more usable proprietary data per operator than almost any other sector in the corpus, and it is publishing surveys about itself instead.

Where the data is the product, the pattern reverses

Real estate and automotive have the corpus's highest own-data ratios: three of five and three of four. Both are sectors where the publisher's business *is* a dataset: listings, registrations, transactions, fleet costs. When the data is already the product, publishing an index is a small step rather than a new capability.

The lesson generalises past those two sectors. The organisations publishing own-data research in this corpus are rarely the ones with the biggest research budgets. They are the ones for whom the dataset already existed and someone decided to point it outward.

11 The six formats as a playbook

Each format costs something different to produce and returns something different. The pickup column is the median distinct Greek outlets covering a release, from the sample of 25 programmes measured in section 5.

WHAT EACH FORMAT COSTS AND WHAT IT RETURNS

FORMAT	IN CORPUS	MEAN PICKUP	COSTS	BEST FOR
Teaser over paid data	6	6.0	Nothing extra: the dataset is already sold	Data businesses. Highest pickup in the corpus
Footprint study	4	5.3	Highest. An independent institute, commissioned	Licence to operate. Aimed at government, not buyers
Own operational data	18	4.6	Analyst time. No fieldwork	Category ownership. Cannot be replicated
Commissioned survey	26	2.7	Panel fieldwork, recurring	Markets where you hold no data of your own
Desk analysis	11	n/a	Analyst time only	Institutions with standing economics teams
Deep whitepaper	3	0.0	High. Long-form writing and design	The reader who already has the problem. Not a press play

Read the first and last rows together. The format with the highest media pickup costs almost nothing extra to produce, because the dataset is already being sold. The format with the lowest, zero, in both sampled cases, is the most expensive to write.

The cheapest format travelled furthest. The most expensive did not travel at all.

That is not an argument against deep whitepapers. It is an argument against expecting one to do a press release's job. The two formats answer different questions: one earns coverage, the other earns a meeting.

The corpus's own investment pattern runs against this. Commissioned surveys are the most common format at 26 of 68, and they sit fourth of five on measured pickup.

A Appendix A: the census

All 68 included publication programmes. Two further programmes were coded and excluded under the published inclusion rules and are retained in the dataset with their exclusion reason.

THE FULL CORPUS

PUBLISHER	SECTOR	PROGRAMME	FORMAT	CADENCE	ACCESS
Hosthub	Tourism and hospitality	The Top 100 Issues Hosts Have to Handle	deep whitepaper	one-off	email- gated
Hosthub	Tourism and hospitality	Annual Greek STR statistics	own data	annual	ungated
Hosthub	Tourism and hospitality	STR share of Greek GDP research	own data	one-off	ungated
Grant Thornton Greece	Tourism and hospitality	Hotels in Greece / hospitality studies	commissioned survey	annual	ungated
Deloitte Greece	Tourism and hospitality	The Greek Hospitality Reimagined	deep whitepaper	one-off	ungated
Deloitte Greece x Hellenic Hoteliers Federation	Tourism and hospitality	The Future of the Greek Hotel	commissioned survey	one-off	ungated
GBR Consulting	Tourism and hospitality	Hospitality Newsletter	teaser over paid	quarterly	ungated- teaser
ITEP x Hellenic Chamber of Hotels	Tourism and hospitality	Annual Survey of the Hotel Sector	commissioned survey	annual	ungated
INSETÉ	Tourism and hospitality	Greek tourism annual report and regional studies	own data	annual	ungated
INSETÉ x Deloitte-Remaco	Tourism and hospitality	Greek Tourism 2030 Action Plans	desk analysis	one-off	ungated
SETÉ / Marketing Greece	Tourism and hospitality	Destination and inbound-visitor research	desk analysis	periodic	ungated
XRTC Business Consultants	Shipping	Annual Report on the Greek Ferry Market	commissioned survey	annual	ungated
Petrofin Research	Shipping	Petrofin Bank Research and Index of Greek Ship Finance	own data	annual	ungated
Moore Greece	Shipping	Moore Maritime Index	own data	annual	ungated
IOBE x Athenian Brewery	Food and beverage	The Brewing Sector in Greece	footprint study	annual	ungated
IOBE x Coca-Cola HBC	Food and beverage	Socio-economic footprint incl. HORECA	footprint study	periodic	ungated
Coca-Cola Greece x AUEB	Food and beverage	Sustainability in HORECA - what consumers want	commissioned survey	one-off	ungated
IOBE x SEVT	Food and beverage	The food and beverage industry in Greece	footprint study	annual	ungated
IELKA	E-commerce and retail	Retail and consumer research programme	commissioned survey	continuous	ungated

PUBLISHER	SECTOR	PROGRAMME	FORMAT	CADENCE	ACCESS
NielsenIQ Greece	E-commerce and retail	FMCG market data releases	teaser over paid	periodic	ungated-teaser
Circana Greece	E-commerce and retail	Consumer trends and FMCG releases	teaser over paid	periodic	ungated-teaser
IOBE x SFEE	Pharma and health	The pharmaceutical market in Greece - Facts and Figures	footprint study	annual	ungated
IQVIA Greece	Pharma and health	ePharmacy audit insights	teaser over paid	periodic	ungated-teaser
Convert Group	E-commerce and retail	Greek ePharmacy / eGrocery reports	teaser over paid	annual+quarterly	teaser-plus-paid
GR.EC.A x ELTRUN AUEB	E-commerce and retail	Annual Greek E-Commerce Survey	commissioned survey	annual	ungated
GR.EC.A x Truberries	E-commerce and retail	E-shop and courier B2B market research	commissioned survey	one-off	ungated
GR.EC.A x SySPaL University of the Aegean	E-commerce and retail	Greek logistics sector study	commissioned survey	one-off	ungated
Sk routz	E-commerce and retail	Annual purchasing behaviour and merchant review	own data	annual	ungated
Nexi Greece	Payments and fintech	Ecommerce Report Greece	own data	annual	ungated
SEAA	Automotive and mobility	Vehicle registration statistics	own data	monthly	ungated
Car.gr	Automotive and mobility	Annual search and market infographics	own data	annual	ungated
Ayvens Greece	Automotive and mobility	Car Cost Index - Greek cut	own data	annual	ungated
INTERAMERICAN and Anytime x NTUA	Automotive and mobility	Road safety and driving behaviour research	commissioned survey	periodic	ungated
Eurobank	Banking, finance and insurance	Economic Analysis and Research	desk analysis	weekly+studies	ungated
Alpha Bank	Banking, finance and insurance	Economic Research bulletins	desk analysis	weekly	ungated
National Bank of Greece	Banking, finance and insurance	Sectoral studies and economic analyses	desk analysis	periodic	ungated
Piraeus Bank	Banking, finance and insurance	Epi Gis agricultural economy periodical	desk analysis	periodic	ungated
Piraeus Financial Holdings	Banking, finance and insurance	Sectoral studies / economic analysis	desk analysis	periodic	ungated

PUBLISHER	SECTOR	PROGRAMME	FORMAT	CADENCE	ACCESS
PwC Greece	Professional services	Greek CEO Survey and publications	commissioned survey	annual	ungated
EY Greece	Professional services	EY Attractiveness Survey Greece	commissioned survey	annual	ungated
Deloitte Greece	Professional services	Greek CFO Survey	commissioned survey	annual	ungated
KPMG Greece	Professional services	Sector surveys - real estate, transport and logistics	commissioned survey	periodic	ungated
ICAP CRIF	Professional services	Greek sector studies	teaser over paid	continuous	teaser-plus-paid
EAAE	Banking, finance and insurance	Annual statistical report of the insurance market	own data	annual	ungated
Spitogatos	Real estate	Spitogatos Property Index	own data	quarterly	ungated
Danos / BNP Paribas Real Estate	Real estate	Greek market reports	own data	quarterly	ungated
Cushman and Wakefield Proprius	Real estate	Greece Marketbeat snapshots	own data	quarterly	ungated
Algean Property	Real estate	Research and yields reports	desk analysis	annual	ungated
American-Hellenic Chamber of Commerce	Real estate	Property Market Outlook for Greece	desk analysis	biannual	ungated
Workable	HR and labour	Free HR reports, ebooks and Hiring Pulse	deep whitepaper	monthly+annual	email-gated
Epignosis / TalentLMS	HR and labour	Annual L&D Benchmark Report	commissioned survey	annual	ungated
Randstad Greece	HR and labour	Employer Brand Research Greece	commissioned survey	annual	ungated
ManpowerGroup Greece	HR and labour	Employment Outlook Survey	commissioned survey	quarterly	ungated
kariera.gr x AUEB	HR and labour	People Pay and AI research	commissioned survey	annual	ungated
Obrela Security Industries	Technology and cyber	Digital Universe Report	own data	annual	ungated
Endeavor Greece	Startups and venture capital	Greek tech ecosystem reports and Year in Review	desk analysis	annual	ungated
Marathon Venture Capital	Startups and venture capital	Greek Startup Compensation Report	commissioned survey	annual	ungated
Marathon Venture Capital	Startups and venture capital	Greek startup funding rounds and exits	own data	annual	ungated

PUBLISHER	SECTOR	PROGRAMME	FORMAT	CADENCE	ACCESS
Found.ation x EIT Digital	Startups and venture capital	Startups in Greece annual report	commissioned survey	annual	ungated
IENE	Energy	The Greek Energy Sector annual report	desk analysis	annual	ungated
SEV	Cross-sector bodies	O Sfygmos tou Epicheirein annual business survey	commissioned survey	annual	ungated
IME GSEVEE	Cross-sector bodies	Biannual SME economic climate survey	commissioned survey	biannual	ungated
diaNEOsis	Cross-sector bodies	Research programme	commissioned survey	continuous	ungated
Focus Bari	Cross-sector bodies	Public opinion and media research releases	commissioned survey	periodic	ungated
Metron Analysis	Cross-sector bodies	Metron Forum polls	commissioned survey	monthly	ungated
Athens International Airport	Tourism and hospitality	Passenger traffic statistics	own data	monthly	ungated
Wolt Greece	E-commerce and retail	Consumer Report	own data	annual	ungated
COSMOTE x ELTRUN AUEB	Technology and cyber	Digital Readiness of Greek SMEs	commissioned survey	biennial	ungated

c **Appendix C: every search, including the empty ones**

All 36 queries run to build the corpus, with what each returned. **11 returned nothing.** They are printed because a census that shows only its successful searches is not reproducible, and because the empty rows are where the white space in section 6 comes from.

SEARCH LOG: HIGHLIGHTED ROWS FOUND NOTHING

LANG	QUERY	FOUND
EN	Petrofin Research Greek shipping annual report free bank research index	1
EN	Marathon VC Found.ation Greek startup ecosystem annual report free download	3
EL	ΣΦΕΕ IOBE μελέτη φαρμακευτική αγορά Ελλάδα δωρεάν έκθεση	1
EL	Τράπεζα Πειραιώς κλαδικές μελέτες αγροτικός τομέας δωρεάν pdf	2
EL	IENE ΔΕΠΑ ΔΕΗ ετήσια έκθεση ενέργεια Ελλάδα μελέτη δωρεάν	1
EL	ΕΑΕΕ ασφαλιστική αγορά Ελλάδα ετήσια έκθεση στατιστικά δωρεάν	1
EL	Algean Property Prosperity Colliers Greece market report ακίνητα έρευνα δωρεάν	1
EL	Kariera.gr Skywalker έρευνα αγορά εργασίας Ελλάδα report δωρεάν μισθοί	2
EL	COSMOTE Nova Vodafone Ελλάδα έρευνα ψηφιακές συνήθειες report δωρεάν μελέτη	0
EL	Linkwise Generation Y ελληνικό digital marketing report έρευνα δωρεάν affiliate	0
EL	ΕΛΤΑ Courier Speedex Geniki Tachydromiki έρευνα ecommerce delivery Ελλάδα report	0
EL	Nielsen Circana Ελλάδα λιανική έρευνα καταναλωτές δωρεάν report FMCG	2
EN	TalentLMS Epignosis research report free download training survey	1
EN	Moore Greece XRTC shipping survey Greek ship finance annual report free	2
EL	ΣΕΒ έρευνα μελέτη επιχειρήσεις Ελλάδα δωρεάν special report	1
EL	Εθνική Τράπεζα κλαδική μελέτη οικονομική ανάλυση δωρεάν pdf sectoral report	1
EL	ΓΣΕΒΕΕ εξαμηνιαία έρευνα μικρομεσαίες επιχειρήσεις Ελλάδα δωρεάν αποτελέσματα	1
EL	diaNEOsis έρευνες μελέτες Ελλάδα δωρεάν χρηματοδότηση επιχειρήσεις	1
EL	TITAN Μυτιληναίος Metlen IOBE μελέτη κοινωνικοοικονομικό αποτύπωμα δωρεάν	0
EL	Public Πλαίσιο Κωτσόβολος έρευνα καταναλωτών τεχνολογία Ελλάδα report δωρεάν	0
EL	Focus Bari MRB Metron Analysis Kara Research δωρεάν έρευνα αποτελέσματα Ελλάδα	2
EL	IQVIA Ελλάδα φαρμακείο αγορά έρευνα report δωρεάν στοιχεία	1
EN	Colliers Greece Cushman Wakefield Proprius Athens market report free download real estate	1
EL	ΣΕΒΤ τρόφιμα ποτά ελληνική βιομηχανία ετήσια έκθεση κλάδου δωρεάν	1
EL	ΟΠΑΠ IOBE μελέτη τυχερά παιχνίδια Ελλάδα αποτύπωμα δωρεάν έκθεση	0
EL	Viva.com έρευνα πληρωμές μικρές επιχειρήσεις Ελλάδα report δωρεάν	0

LANG	QUERY	FOUND
EL	Enterprise Greece Athens Exchange EBEA έρευνα επενδύσεις Ελλάδα έκθεση δωρεάν	0
EL	Odyssey Cybersecurity Encode Ελλάδα threat report δωρεάν λήψη	0
EL	Marketing Greece ΣΕΤΕ έρευνα τουριστικό brand Ελλάδα μελέτη δωρεάν	1
EL	Deloitte Greece CFO Survey EY Greece CEO Outlook ελληνικές επιχειρήσεις έρευνα δωρεάν	2
EL	Ελληνο-Αμερικανικό Εμπορικό Επιμελητήριο έρευνα μελέτη οικονομία δωρεάν έκδοση	1
EL	Διεθνής Αερολιμένας Αθηνών ΟΛΠ ετήσια έκθεση κίνησης στατιστικά δωρεάν δεδομένα	1
EL	HELLENiQ ENERGY Motor Oil ΔΕΗ IOBE μελέτη αποτύπωμα ενέργεια Ελλάδα έκθεση	0
EL	Worldline Cardlink Ελλάδα δεδομένα συναλλαγών έρευνα report δωρεάν καταναλωτική δαπάνη	0
EL	efood Wolt Ελλάδα ετήσια στοιχεία παραγγελίες report δεδομένα δημοσίευση	1
EL	Quest Group Info Quest COSMOTE έρευνα ψηφιακή ωριμότητα ελληνικές επιχειρήσεις report	1

B Appendix B: how everything was coded

Every row in Appendix A was coded on fourteen variables. The rules below were written when each ambiguity was first hit, not reconstructed afterwards, so any row can be re-derived or disputed.

The unit

The unit is the **publication programme**, a publisher plus a recurring title, not the organisation and not the individual edition. Cadence, format, gating and method disclosure are properties of a programme, not of a company. One consultancy in this corpus runs three programmes with different formats and audiences; collapsing them to a single row would lose exactly the variation the paper is about. Both counts are reported: **68 programmes across 64 publishers**.

Inclusion

All three had to hold: the publisher is Greek or substantively Greek-operating; the research is **original**: own data, own survey, or a commissioned study, not a translated global edition; and it is **free** in whole or as a substantive teaser, where a teaser must carry a finding rather than a table of contents.

Exclusions, and why

CASE	DECISION	REASON
State bodies and regulators	Excluded	Statutory publication is not the behaviour being studied
Universities publishing alone	Excluded	Included only as a co-signer on company- or association-funded work
Business associations	Included	Company-funded bodies publishing on members' behalf: the same behaviour with the cost pooled
Free software tools with dashboards	Excluded	A tool is not a publication
Figures reaching the press only via filed accounts	Excluded	Journalism about a company is not publication by it

The six formats

A own operational data · **B** deep practical whitepaper · **C** commissioned survey · **D** institute-authored footprint study · **E** free teaser over a paid dataset · **F** desk or analytical study.

Where a programme could be two, the **data basis decides**: own operational data is A, new primary collection is C, neither is F. If the full dataset is sold, E overrides everything. Code D requires an **independent author**: a self-authored footprint claim is A or F with a policy motive, not D.

The motive decision rule

Motive is inferred from four observable signals, applied in order, and never asserted:

SIGNAL	READING
Author	Independent institute or university on the cover → licence to operate or policy. Self-authored → commercial
Gate	Email-gated → demand generation, always
Headline metric	Jobs, GDP share, multipliers → licence to operate. Behaviour, prices, indices → category ownership. Practical how-to → demand generation
Addressee	Government, media, society → policy. Buyers → demand generation or category ownership

Ties resolve by addressee. Where the publisher is an association and the addressee is the state, the code is **policy** even where the work also serves members commercially. **CSR framing is coded separately from motive**, so a reader can re-derive the CSR answer without trusting the motive column.

The full code list. Eight codes are in use. An earlier version of this appendix defined only six, which the round-three reliability check caught:

CODE	DEFINITION
category-ownership	Owning the recurring number in a category
demand-gen	Generating or qualifying commercial demand
policy	Addressed to the state or to public debate
teaser-to-paid	Free findings existing to sell the dataset
pr-amplification	A press release rather than a document; no deeper purpose visible
licence-to-operate	Defending the publisher's right to operate
csr-marketing	Contribution or responsibility framing used as brand marketing rather than to defend a licence. Four programmes
employer-brand	Addressed to candidates rather than buyers or the state. One programme

csr_framed takes three values, not two: yes, no, and partial where the artefact carries contribution framing in part of the document only. Two programmes are partial.

Reliability, and the round that failed

Coding reliability was checked three times, twice by re-coding a systematic 20% sample blind against the same source artefacts, and once by an independent second coder against the whole corpus.

Round one failed at 0.64 against a 0.80 threshold. It produced three rule corrections and one schema change: desk analyses had no code and were defaulting into commissioned surveys, putting eleven programmes in the wrong bucket. **Round two, on a fresh sample drawn with a different start after recoding, reached 0.93.**

Both of those rounds were **intra-rater**: one coder re-checking their own work at a distance. That catches inconsistent application of a rule. It does not catch a rule that is wrong in a way one person repeats consistently.

Round three: independent coding, and what it broke

On 6 August 2026 the whole corpus was coded a second time by an independent agent working only from the published rules and the same visible facts about each programme, blind to the original codes. Not a sample: **all 68 programmes**. The second coder was a large language model, which is a weaker instrument than a trained human second coder and is stated as such below. What it tests precisely is whether *the published rule*, applied by somebody who was not there when it was written, reproduces the decision.

7
INTER-RATER AGREEMENT, N=68, ALL THREE JUDGEMENT COLUMNS. THRESHOLD FOR THIS PROJECT IS K 0.80.

COLUMN	RAW AGREEMENT	COHEN'S K	VERDICT
format	85.3%	0.80	Meets the threshold
csr_framed	89.7%	0.34	High agreement, unstable κ: see below
motive_primary	51.5%	0.34	Fails

The format schema holds. The motive column does not. Two coders working from the same published rule agreed on the format of a programme 85% of the time, and on why it was published just over half the time. That is the single most important thing this paper learned about itself, and it lands on the column that answers its own headline question.

Three things caused it, and only two are fixable.

One: the published rule was incomplete. The corpus uses eight motive codes. The decision rule printed above listed six: `csr-marketing` and `employer-brand` appear in the results table but were never defined, so five programmes could not possibly have been matched. The same defect applied to `csr_framed`, which the rule described as yes/no while the data carries a third value, partial, in two rows. Both are now documented above. Excluding those five rows moves κ to 0.38: it was not the main cause.

Two: the gate signal was read as an equivalence. Eleven of the 34 motive disagreements are the same disagreement: programmes I coded demand-generation, the second coder coded category-ownership, and *all eleven are ungated*. The rule says email-gated implies demand-generation. It does not say the reverse, but it does not say what ungated implies either, so the decision falls through to signals three and four, and for a professional-services firm's free sector report, “practical how-to” and “benchmark” are genuinely both true. The rule underdetermines the answer for a whole class of publisher.

Three: the column may not be reliably codeable from artefacts at all. Motive is inferred, and this result is the measurement of how much that inference costs.

What follows from it. The motive distribution stays in the paper, because deleting an inconvenient result after measuring it is worse than publishing it with its reliability attached. But it is now labelled: **every motive figure in this paper is a single-rater judgement with κ 0.34, and should be read as indicative, not measured**. The corpus was *not* recoded to chase a better number: recoding after seeing the disagreements is how a reliability check becomes theatre.

The CSR answer survives this intact, and that is by design rather than luck. `csr_framed` was coded independently of motive from the beginning precisely so the headline question would not depend on the softest column. Its raw agreement is 89.7%. Its κ is low for a well-understood reason: 59 of 68 rows carry the same value, so chance agreement is already 84.5% and κ becomes unstable: the arithmetic penalises a skewed column even when the coders agree. On a question with a lopsided true answer, raw agreement is the more honest statistic, and it is reported alongside.

Second coder: DeepSeek deepseek-chat, temperature 0, one pass, no retries on disagreement. Prompt carried the published rules verbatim plus nine observable fields per programme; the original codes were withheld. Per-row output in `inter-rater-check.csv`.

Link integrity

All 70 coded rows carry a source URL. At the most recent check, **66 resolved live, four returned 403 to an automated client** and were reached by other means, and **none were broken**. Two URLs found broken in an earlier check were corrected rather than dropped.

12 References and instruments

This paper rests almost entirely on primary research conducted for it. That makes the instruments themselves the references, and all of them are published in full at the URLs below. The two external statistics an earlier

draft relied on were both cut at citation audit, and they are listed here rather than deleted from the record.

Instruments and datasets, all published with this paper

- 1 **Broikos, N. (2026). Census of Greek corporate research publishing.** Dataset. 70 publication programmes from 65 publishers, coded on 20 variables; 68 included and 2 excluded with their exclusion reason recorded. Coded 5-6 August 2026. Every row carries the source URL of the artefact coded.
<https://broikos.gr/research/data/paper-01/census.csv>
This file is the answer to any question about an individual publisher in this paper. The 68 organisations analysed are not listed separately in these references because each is a row in this dataset with its own URL.
- 2 **Broikos, N. (2026). Search log.** Dataset. All 36 documented searches in Greek and English, including the 11 that returned nothing.
<https://broikos.gr/research/data/paper-01/search-log.csv>
Published so that the corpus boundary can be audited. Every count in this paper is a floor, and this file is the evidence for how the floor was reached.
- 3 **Broikos, N. (2026). Coding rules.** Full rule set, with each ambiguity recorded at the point it was first encountered rather than reconstructed afterwards. Includes the eight motive codes and the three values of the CSR-framing variable.
<https://broikos.gr/research/data/paper-01/coding-rules.md>
- 4 **Broikos, N. (2026). Media amplification measurement (M1).** Dataset. 25 programmes stratified across formats; one dated search each; distinct Greek online outlets counted. Measured 5-6 August 2026.
<https://broikos.gr/research/data/paper-01/amplification.csv>
Reported throughout as a floor. Censored at the search-result ceiling in 7 of 25 cases; online press only.
- 5 **Broikos, N. (2026). Answer-engine citation test (M2): locked prompt set.** Twelve prompts in Greek, committed and dated 5 August 2026, before any engine was queried.
<https://broikos.gr/research/data/paper-01/prompts-m2.md>
Pre-registration of the protocol. The file also carries the run record and the deviation from design: three engines were run, not the four specified, because no Microsoft Copilot API access was available.
- 6 **Broikos, N. (2026). Answer-engine citation log (M2).** Dataset. 36 responses, 303 source citations, coded by domain against the census. Run 6 August 2026.
<https://broikos.gr/research/data/paper-01/ai-citations.csv>
Engines and versions as queried: OpenAI gpt-5-search-api via the chat-completions endpoint; Perplexity sonar; Google gemini-2.5-flash with Google Search grounding. One session per engine, first response only, no follow-ups, run from Greece. One Gemini response returned an API error and is counted as an answer naming no sources.
- 7 **Broikos, N. (2026). Coding reliability records.** Three rounds: two intra-rater test-retest on systematic 20% samples, and one inter-rater round across all 68 rows.
<https://broikos.gr/research/data/paper-01/reliability-check.csv>
<https://broikos.gr/research/data/paper-01/reliability-check-2.csv>
<https://broikos.gr/research/data/paper-01/inter-rater-check.csv>
Round 1 failed at 0.64 and is published in full, including the three rule corrections and the schema change it forced. Round 3 used DeepSeek deepseek-chat at temperature 0 as an independent second coder, blind to the original codes, and is published row by row including every disagreement. It failed on the motive column at k 0.34.

- 8 **Broikos, N. (2026). International comparison corpus.** Dataset. 119 publication programmes across Greece, Europe and America on a matched schema.
<https://broikos.gr/research/data/paper-01/international-corpus-119.csv>
***Lower evidential standard than the census, and used only for distributional comparison.** 64 of the 119 rows are confirmed and 55 unverified; no reliability check and no second coder. Stated as such wherever it is used.*
- 9 **Broikos, N. (2026). Claim ledger and source log.** The verification status of every figure considered for this paper.
<https://broikos.gr/research/data/paper-01/claims-claim-ledger.csv>
<https://broikos.gr/research/data/paper-01/source-source-log.csv>

Sources considered and cut

Both external statistics an earlier draft used to argue that published research works commercially were removed at citation audit. They are recorded here because a paper that grades other people's sourcing owes the reader its own casualties.

- 1 **Edelman and LinkedIn. 2025 B2B Thought Leadership Impact Report.**
<https://www.edelman.com/expertise/Business-Marketing/2025-b2b-thought-leadership-report>
***CUT.** Could not be opened at source within the verification window. The finding may well be sound; the citation was not verifiable, so it was removed rather than carried on the strength of secondary reporting.*
- 2 **Demand Gen Report. Content Preferences Benchmark Survey.**
<https://www.demandgenreport.com/resources/content-preferences-survey/>
***CUT: misattribution.** The widely-quoted claim that only 5% of buyers are in-market at any moment was traced and found not to originate with the report it had been attributed to, but with separate research quoted inside it. This is failure mode F2 in the citation taxonomy used for this project: attributing a finding to the document that repeated it rather than to the one that produced it.*

Coded exemplars, with their URLs printed

The census codes 68 publication programmes from their published artefacts. Naming all 68 in the text would make this a directory rather than a paper, and the full coded dataset is published separately. **These seven are the exemplars the text discusses by name**, listed here with live URLs so a reader can check the coding against the artefact rather than taking it on trust. All seven were re-checked on 7 August 2026.

- 1 **Hosthub. The Top 100 Issues Hosts Have to Handle (Greek edition).** Own-data format exemplar; more than a million guest messages; email-gated.
<https://www.hosthub.com/el/blog/whitepaper-ta-100-korifaia-zitimata/>
- 2 **Hosthub. Short-term rentals in Greece.** Second programme from the same publisher, used for the edition-count coding.
<https://www.hosthub.com/blog/str-in-greece/>
- 3 **Convert Group. Greek ePharmacy report, 2024.** Own-data exemplar built on transaction panel data.
https://convertgroup.com/reports_posts/greek-epharmacy-2024/
- 4 **Randstad Greece. Press releases and research index.** Commissioned-survey exemplar with a continuous edition series.
<https://www.randstad.gr/h-randstad/deltia-typou/>
- 5 **Obrela Security Industries. Reports index.** Own-operational-data exemplar in a technical sector.
<https://www.obrela.com/resources/reports>
- 6 **Workable. Free ebooks and reports.** International comparator for the format coding.
<https://resources.workable.com/free-ebooks-and-reports>

7 **Broikos, N. Prior whitepaper research dataset.** The 119-programme lighter scan used for the international comparison in section 06.

<https://broikos.gr/research/data/paper-01/>

A note on what is not cited here

There is no literature review in this paper and no academic citation apparatus, and that is a deliberate limitation rather than an oversight. The paper makes an empirical claim about a population of documents in one country and defends it with a census, not with a theoretical framework. A reader looking for the signalling-theory or content-marketing literature that would situate this work will not find it, and should treat the absence as a boundary on what the paper claims.

13 Citation, licence and interests

HOW TO CITE THIS PAPER

Broikos, N. (2026) *Who publishes the numbers: a census of Greek corporate research publishing*. Version 1.0, 6 August 2026. Athens. Available at:
<https://broikos.gr/research/who-publishes-the-numbers.pdf> (Accessed: DD Month YYYY).

VERSION

Version 1.0, published 6 August 2026. This is the first public version. Substantive changes will increment the version number and be listed in the change log published with the data.

DATA AVAILABILITY

Everything underlying this paper is published at <https://broikos.gr/research/data/paper-01/> under the same licence as the text. This is the full working apparatus, not a summary extract:

census.csv: all 70 coded programmes with 20 variables each, including the two excluded rows and their exclusion reason, and the source URL for every row. search-log.csv: all 36 documented searches including the 11 that found nothing. coding-rules.md: the full rule set with every ambiguity recorded at the point it was hit. amplification.csv: the media pickup measurement. reliability-check.csv and reliability-check-2.csv: the two intra-rater rounds, including the round that failed at 0.64. inter-rater-check.csv: the round-three independent coding, row by row, including every disagreement. ai-citations.csv and prompts-m2.md: the answer-engine test: the prompts as locked before the run, and every source named in all 36 responses. claim-ledger.csv and source-log.csv.

The paper is designed so that a reader who disagrees with a coding decision can find the row, read the rule, and re-derive it. That is the point of publishing the failed reliability round and the disagreement log rather than only the final figures.

No confidential data was used. Every coded programme is a publicly available document.

DECLARATION OF INTERESTS

The author produces corporate research and would sell research services. This paper argues that Greek companies under-use a capability the author sells, which is a conflict of interest of the most direct kind and is stated here as plainly as it can be.

Three specific disclosures. The author's own e-commerce businesses are not in the census and were not eligible for it. No organisation coded here was contacted, consulted or given sight of the paper before publication, so no coding decision could have been influenced by a commercial relationship. And the author's own coding failed its first reliability check at 0.64 and its motive column failed the independent check at κ 0.34; both are printed in full rather than repaired quietly, which is the only real defence available against the interest declared above.

LICENCE

This paper and the datasets published with it are released under a **Creative Commons Attribution 4.0 International licence (CC BY 4.0)**. You may copy, redistribute, quote, chart and build on this material, including commercially, provided you credit the source. Licence text:

<https://creativecommons.org/licenses/by/4.0/>, and served beside this paper at

<https://broikos.gr/research/LICENSE-CC-BY-4.0.txt>. What the grant covers and what it withholds is itemised in <https://broikos.gr/research/LICENSING.md>. Analysis code is MIT:

<https://broikos.gr/research/LICENSE-MIT.txt>

FUNDING AND INDEPENDENCE

Unfunded. No sponsor, client, trade body or commissioning party paid for, commissioned, reviewed or approved this research. No organisation named in it was given sight of it before publication. This matters because a substantial share of the sources graded here *were* commissioned, and the paper argues that a commissioning interest should be stated wherever it exists.

CORRECTIONS

If a figure here is wrong, or an organisation is described incorrectly, write and it will be corrected with the date and the reason recorded in a public change log rather than silently edited. That log is published at <https://broikos.gr/research/corrections.md> and already records every figure cut, corrected or downgraded during production. Every correction made during production is already recorded in the claim ledger published with this paper, including the figures that were cut.

Contact: <https://broikos.gr/contact> · <https://broikos.gr>

ABOUT THE AUTHOR

Nikolaos Broikos operates e-commerce businesses in Greece, works in web development and digital strategy, and builds agent harnesses. He writes from Athens.

This programme is the record of that work rather than a commentary on it. The e-commerce operations supply the transaction data behind the landed-cost model in *What a Greek online order actually costs*; the harnesses are the instruments measured in *What actually makes an agent harness work*. Where the author's own systems are the subject, that is stated in the first paragraph of the paper concerned as well as in its declaration of interests.

No academic affiliation, no institutional backing, no funding, and no client commissioned any of this.

The papers therefore ask to be judged on their published instruments, data and corrections rather than on credentials: every dataset, every claim ledger including the claims that failed, and every retraction is published alongside the text, so a reader who distrusts the author can check the work instead.

THE OTHER PAPERS IN THIS PROGRAMME

Independent, unfunded and published free under CC BY 4.0, each with its underlying data. Read separately; they share a method, not an argument.

1 What a Greek online order actually costs

<https://broikos.gr/research/what-a-greek-online-order-costs.pdf>

What the €36.1bn e-commerce figure counts and what it does not, the convergence of Greek and European online buying, and a landed-cost model built from published tariffs rather than quoted rates.

2 The adoption gap: what businesses say about AI and what the statistics measure

<https://broikos.gr/research/the-adoption-gap.pdf>

Where Greek firms actually sit on AI adoption once the size classes are separated, why the skills-gap explanation does not survive the spending data, and what Greek firms bought instead.

3 What actually makes an agent harness work

<https://broikos.gr/research/agent-harness.pdf>

Techniques, prompt engineering and measured results from three agent harnesses, two built by the author and one not, across 45 graded sessions on a single model, including three faults found in the measuring apparatus itself.

4 The Greek e-shop technical audit

<https://broikos.gr/research/greek-eshop-audit.pdf>

900 Greek domains measured on one day: what Greek online retail actually sends to a browser. 36.8% send no security headers, 11.1% fail a validating TLS handshake, and 5 shops out of 505 pass five elementary checks.

5 The Greek SME digital bill

<https://broikos.gr/research/greek-sme-digital-bill.pdf>

What it costs per year to sell online in Greece, priced entirely from published pages: EUR 157.50 at the floor. Of 25 cost lines, 7 have no published price at all, and they are disproportionately the mandatory ones.

6 What an agent actually costs

<https://broikos.gr/research/what-an-agent-costs.pdf>

Seven versions of one production agent on the same benchmark and the same model, including the two versions that got worse, the failure taxonomy, and the chart that would have shown a cost explosion that never happened.

ABOUT THE DATA

68 included publication programmes, coded 5-6 August 2026 from 36 logged searches in Greek and English. Two further programmes were coded and excluded under the published inclusion rules and are retained in the dataset with their exclusion reason.

Media amplification measured 5-6 August 2026 across a stratified sample of 25 programmes. The answer-engine citation test was run on 6 August 2026 against three engines, with all 36 responses published.

RELIABILITY

Three rounds. Two intra-rater: round 1 failed at 0.64 and produced three rule corrections and a schema change; round 2, on a fresh sample after recoding, reached 0.93. Round 3 was inter-rater across all 68 rows: format κ 0.80, motive_primary κ 0.34.

Of 70 source URLs checked, 64 resolved live, 4 blocked automated clients, and 2 were broken and corrected.

CORRECTIONS

Every organisation here is coded on variables inferred from published artefacts, including motive. If you publish research in Greece and I have coded you incorrectly, or missed you entirely, tell me and I will fix it with the date and reason recorded.

broikos.gr

© 2026 Nikolaos Broikos. The census, search log, coding rules, inter-rater record and answer-engine log are published alongside this paper so any row can be checked, disputed or extended.