



REPORT • AUGUST 2026

Master data in Polish enterprises

*Market, regulation and the practice of getting customer,
material and employee data in order*



SNOK Sp. z o.o. • Warsaw • snok.ai

Contents

1.	Executive summary	3
2.	A year of consolidation in the data management market	5
3.	Polish enterprises and their data: the starting point	7
4.	KSeF: a nationwide test of customer master data	11
5.	The CIO calendar: migrations that will not wait	14
6.	AI starts with the master file	18
7.	Why data quality programmes fail	20
8.	The incremental approach: method and metrics	23
9.	Conclusions	26
10.	Methodology and sources	27
	About SNOK	28

1

Executive summary

This report organises what is known about master data: the records describing customers, vendors, materials, products, employees and financial accounts. The figures it draws on were published between 2020 and 2026; each is quoted with its source and date.

1 The market has consolidated at a pace not seen in years.

Salesforce closed its acquisition of Informatica in November 2025, SAP acquired Reltio in May 2026, and in April 2026 Gartner reinstated its Magic Quadrant for this category after a gap of more than four years. Both acquisitions were announced with the same rationale: making data ready for artificial intelligence.

2 Polish enterprises enter this period with low data maturity.

According to EY, only 9 percent of mid-sized and large companies in Poland have a complete data infrastructure ready to feed AI models (survey fielded at the turn of 2025 and 2026). Just 24.5 percent of Polish firms perform data analytics (Eurostat, 2025), and 8.7 percent use AI technologies, against an EU average of 20 percent (Eurostat, 2025).

3 KSeF, Poland's mandatory e-invoicing system, has become the first nationwide test of customer master data.

Between 1 February and 26 May 2026 the system processed more than 289 million invoices from over 2 million issuers (Ministry of Finance). Crucially, KSeF does not verify that the buyer's tax ID belongs to the intended counterparty; a master file error means an invoice delivered to the wrong company and a correction to zero, because correction notes no longer exist.

4 The end of SAP ECC support is forming a migration queue, and migration does not repair data.

Mainstream maintenance ends on 31 December 2027. According to DSAG, 54 percent of surveyed organisations still run ECC or older Business Suite releases (2026), and according to ISG, 58 percent of SAP migration programmes overrun budget and schedule (2026). Other vendors have announced similar deadlines: Microsoft will end standard support for Dynamics GP at the close of 2029 (announcement of September 2024).

5 Data quality programmes fail because of scope more often than technology.

Gartner predicted that more than 75 percent of master data management programmes would fail to meet business expectations through 2025 (Magic Quadrant, December 2021). The same mistakes recur: too many data domains at once, no data owner with a real mandate, and record merging left to run without human oversight.

6 The alternative to a multi-year programme is an incremental approach.

One data domain, one success criterion agreed before the start, and results measured on the organisation's own data. The scale of the problem can be estimated internally within days; chapter 8 provides four control questions.

The advantage will belong not to whoever buys the tools first, but to whoever first puts in order the layer those tools work on

2

A year of consolidation in the data management market

Between November 2025 and May 2026 the largest software vendors spent billions of dollars on companies whose business is data quality and consistency. The rationale behind both deals was identical and stated openly: without orderly data there is no credible artificial intelligence.

On 18 November 2025 Salesforce closed its acquisition of Informatica, valued at approximately USD 8 billion (Salesforce announcement). On 27 March 2026 SAP announced its intent to acquire Reltio, a master data management platform vendor, and closed the transaction on 7 May 2026; the stated goal was to make data, both from SAP systems and beyond them, ready for AI (news.sap.com).

A third signal came from the analysts. In April 2026 Gartner reinstated the Magic Quadrant for Master Data Management Solutions, whose previous edition had appeared in December 2021; the new edition rated twenty vendors. Mordor Intelligence values the global market for this category at roughly USD 18.2 billion in 2025, while Grand View Research had already put it at USD 19.9 billion in 2023; both firms project double-digit annual growth.

For boards in Poland, the practical meaning of these events is straightforward. A category long treated as technical back office has been repriced by the market as a precondition for automation. Any organisation planning AI deployments, agents or an ERP migration will sooner or later face a question about the state of its own master files.

~USD 8bn

the price Salesforce paid for Informatica

Salesforce, closing announcement, 18 November 2025

3

Polish enterprises and their data: the starting point

Before asking how to put data in order, it is worth establishing the starting point. Public statistics and market research from 2025-2026 paint a consistent picture: Polish companies have digitalised their processes without putting in order the data those processes run on.

3.1. The systems are there; order in the data is not

According to Statistics Poland (GUS), 40.5 percent of Polish enterprises run an ERP system, up 4.5 percentage points in two years, and 25.1 percent use a CRM ("Information Society in Poland 2025", GUS, October 2025). Paid cloud services are bought by 55.3 percent of companies. The infrastructure, in other words, exists and keeps growing.

Working with data is a different story. Only 24.5 percent of Polish firms perform data analytics, compared with 60 percent in Denmark (Eurostat, 2025). AI technologies are used by 8.7 percent of enterprises according to GUS and 8.4 percent according to Eurostat; the EU average is 20 percent, which places Poland second from last, ahead of Romania only. The difference between GUS and Eurostat stems from methodology, which is why we quote both figures side by side.

3.2. Data named as a barrier, explicitly

Behind the averages lies a steep divide: AI is used by 42 percent of large companies, 15.6 percent of mid-sized ones and 6.1 percent of small firms (GUS, 2025). Digitalisation indicators, however, describe the maturity of the environment rather than measuring master data quality directly; that picture comes from the studies discussed below.

An EY survey of roughly 500 Polish companies (CubeResearch, fieldwork in Q4 2025, results published in 2026) delivers the number worth remembering: only 9 percent of mid-sized and large enterprises in Poland have a complete data infrastructure ready to feed AI models, while a further 30 percent hold traditional data only. Data quality was named among the most significant barriers to adoption, alongside difficulties in operationalising strategy and a skills deficit.

A study by Algolytics and SW Research covering more than 700 companies (2025) paints a similar picture: 73 percent of Polish enterprises say they lack the competencies to exploit their data, and 66 percent do not use data in decision-making.

Our project work points to a simple mechanism behind these declarations. The same customer exists in the ERP under its full legal name, in the CRM as a hand-typed shorthand, and in the procurement system with no tax ID at all. Each system is right locally; none is right in full. The cost of this state rarely appears as a budget line, because it disperses into manual balance reconciliations, payment terms clarified over email, and management reports assembled from several versions of the truth.

Five teams paying for the same duplicate

- **Finance**
reconciling the same customer's balances by hand
- **Procurement**
calling a vendor to check whether two records are one firm
- **Warehouse**
accepting a second delivery of a part already in stock
- **Controlling**
assembling one report from three versions of the truth
- **IT**
maintaining a 'temporary' merge script, three years on

None of the five sees the full cost; not one line of it reaches the budget. Source: SNOK.

9%

of mid-sized and large companies in Poland have a complete data infrastructure ready to feed AI models

EY Poland, CubeResearch survey (c. 500 firms, Q4 2025), published 2026

3.3. Mergers and acquisitions: growing exposure

The Polish market recorded 330 merger and acquisition transactions in 2025, following 348 a year earlier (M&A Index Poland, Fordata and Navigator Capital). The largest deal of 2025 was the sale of a 49 percent stake in Santander Bank Polska to Erste Group for approximately PLN 29.5 billion.

Every acquisition means merging customer files, material catalogues and charts of accounts maintained under different rules. No public research measures the scale of this phenomenon in Poland, so we treat deal volume as a measure of exposure rather than proof. Our projects in post-acquisition groups, including a chemical group consolidating the data of five companies across three countries, show a recurring pattern nonetheless: the first weeks of integration are spent establishing which entity uses which catalogue, and how definitions of seemingly identical concepts differ.

4

KSeF: a nationwide test of customer master data

Poland's mandatory National e-Invoicing System is the first mechanism to test the customer master files of every Polish taxpayer at once, every working day, with no allowance for old habits.

4.1. Scale and timeline

KSeF, Poland's National e-Invoicing System, became mandatory in stages: from 1 February 2026 for enterprises with 2024 sales above PLN 200 million, and from 1 April 2026 for all other VAT payers; the smallest businesses join on 1 January 2027 (Ministry of Finance, ksef.podatki.gov.pl). Since 1 February, every taxpayer has also been obliged to receive invoices through the system.

The scale is unprecedented: between 1 February and 26 May 2026, more than 289.5 million e-invoices were issued in KSeF by over 2 million issuers (Ministry of Finance data reported via the Polish Press Agency, May 2026).

4.2. What the system checks, and what it does not

A structured invoice under the FA(3) schema contains several hundred fields; industry analyses speak of more than three hundred, only part of them mandatory (the Ministry of Finance does not publish an official count). The buyer's data, including the NIP tax identification number, name and address, is filled in by the issuer from its own master file.

The system's most consequential property tends to surprise: KSeF validates the file against the schema, but it does not verify that the buyer's NIP belongs to the counterparty the transaction concerns. An invoice carrying a mistyped, outdated or simply wrong NIP will receive a KSeF number and be made available to whichever entity holds that NIP; if the NIP does not exist, the invoice reaches no one. This position was set out by the Director of National Revenue Information (KIS) in an individual tax ruling of 24 March 2026 (ref. 0114-KDIP1-3.4012.44.2026.1.AMA), issued for the facts described in that case.

The consequences of an error have changed as well. Since February 2026 the buyer can no longer fix someone else's mistake with a correction note; a wrong NIP requires the issuer to issue a correcting invoice down to zero and a new invoice. The issuer is responsible for the accuracy of identification data even when the incorrect data came from the counterparty; the consequences of a specific error depend on the circumstances and the applicable law.

4.3. The first six months in practice

Tax advisers documented the recurring problems of the first weeks in April 2026: EU VAT numbers entered in place of the NIP, the same data placed in different fields by different counterparties, and a parallel circulation of documents in which a PDF sent by email travels alongside the KSeF invoice (Prawo.pl, 2 April 2026). According to one software vendor, Exorigo-Upos, 18 percent of correcting invoices retrieved from KSeF contain errors, and around 10 percent of invoices enter the system with key data missing (Bankier.pl, 29 July 2026); these figures come from a single vendor's client base, not public statistics, and should be read as such.

The Ministry of Finance has announced that penalties connected with KSeF obligations are deferred until the end of 2026 (per Prawo.pl, April 2026). The grace period has an end date, though, and the mechanism is here to stay: the quality of the customer master file has ceased to be an internal accounting matter and has become a condition of doing paperwork with every counterparty, one document at a time.

289.5m

e-invoices issued in KSeF between 1 February and 26 May 2026, by more than 2 million issuers

Ministry of Finance data, via PAP, May 2026

5

The CIO calendar: migrations that will not wait

The support calendars of the largest vendors are lining up a queue of system migrations for the second half of the decade. Market experience shows that the most consistently underestimated element of these projects is the data.

5.1. The support calendar: SAP and beyond

SAP maintains mainstream support for SAP Business Suite 7 applications until the end of 2027, with paid extended maintenance available until the end of 2030 (SAP announcement of February 2020, confirmed in current support documentation). For its largest customers SAP has additionally prepared a subscription-based transition option covering 2031-2033 (SAP announcement, August 2025).

Support deadlines are not an SAP peculiarity. Microsoft has announced the end of development and standard support for Dynamics GP at the close of 2029, with security patches until April 2031, pointing to Dynamics 365 Business Central as one of the target paths (September 2024). On the Polish market, where Comarch and other domestic vendors hold strong positions alongside SAP (IDC, 2024), the same moment arrives with a version upgrade or a post-acquisition consolidation. Every such operation is, in practice, a data migration and an examination of master file quality.

5.2. How many companies face a migration

Meanwhile, according to the February 2026 survey of the German-speaking SAP user group DSAG, 54 percent of responding organisations still run SAP ECC or older Business Suite releases (a multiple-choice question), and only 37 percent plan to leave those systems by the end of 2027. In early 2025 Gartner estimated that more than 60 percent of SAP customers remained on ECC with no migration decision taken (via The Register, February 2025). No public figures exist for Poland; the country's largest SAP partner estimates the total number of SAP ERP users in Poland at more than 2,000 companies (All for One, company material, undated).

The arithmetic is simple: a substantial share of these migrations will happen between 2026 and 2030, competing for the same delivery capacity.

5.3. Data migration as a leading risk

ISG's study "The State of SAP Migration" (January 2026, 200 organisations with over 1,000 employees) shows how such programmes actually run: 58 percent of SAP migrations overrun budget and schedule, and only 15 percent finish on time and on budget. According to a Polish review of the study, 24 percent of respondents point to data migration complexity as a leading risk (erp-view.pl, February 2026). The study's authors draw a conclusion that matters for the rest of this report: carrying years of uncleaned master data into a new system undermines the value of analytics and AI after go-live.

5.4. The SAP case: Business Partner and CVI

In S/4HANA the Business Partner model is mandatory: customers and vendors cease to be separate objects, and the transition to the common model is performed by the Customer Vendor Integration mechanism, executed in the source system as an obligatory preparatory step (SAP Note 2265093 and official SAP conversion materials).

CVI's job is to move data into the new structure, not to improve its quality. Three spellings of the same customer will remain three business partners after conversion unless someone merges them first. A record missing a field the new model requires comes back to the team as an exception for manual handling, typically during the cutover window, when the team has the least time to spare. Deciding whether two similar records describe the same entity is, moreover, a business judgement, not a technical one.

Hence the sequence we recommend in projects: measure duplicates and gaps before the conversion, deduplicate with the data owner involved, and only then run CVI. The same work done after go-live costs more, because it happens under production pressure.

A sequence that holds up



Measure and deduplicate before the migration; after go-live the same work happens under production pressure. Source: SNOK.

We apply this sequence regardless of the target system. Measurement takes days and needs no access to the production environment, so it can precede the project decision itself. Deduplication runs on the live master file in parallel with migration preparations and does not block the schedule. The migration then completes the data work instead of starting it - and during cutover the team looks after the system, not after the question of whether two records are the same company.

58%

of SAP migration programmes overrun budget and schedule;
15 percent finish on time and on budget

ISG, The State of SAP Migration (n=200), January 2026

6

AI starts with the master file

Language models and agents do not check whether data is true. They answer from whatever they are given - with the same confidence whether the data is consistent or contradictory.

The mechanism is easy to picture. An agent asked for the balance of a customer that exists in four variants across systems will answer instantly; it will not flag that it picked one of four possible records. Automation built on inconsistent data does not remove the disorder. It increases its speed and its reach.

Analyst forecasts have been saying the same thing for two years. Gartner expects organisations to abandon 60 percent of AI projects that lack AI-ready data through 2026, and notes that 63 percent of organisations either lack suitable data management practices or are unsure of them (press release, 26 February 2025).

The Polish figures fit that forecast without any stretching. If 8.7 percent of enterprises use AI technologies (GUS, 2025) while 9 percent of mid-sized and large companies report an AI-ready data infrastructure (EY, 2026) - two studies of different populations, but pointing the same way - it is hard to argue that the barrier lies in access to models. The models are available off the shelf; prepared data is not.

The master data layer plays a role in AI architecture that neither a warehouse nor an analytics platform can take over: it decides which record is correct and who is accountable for keeping it that way. Warehouses and lakehouses collect and model data, but they do not rule on its truth; without a golden-record layer, analytics consolidates inconsistencies instead of removing them.

60%

of AI projects will be abandoned through 2026 if they are not supported by AI-ready data

Gartner, press release, 26 February 2025



7

Why data quality programmes fail

The paradox of this field is that the failure statistics of data management programmes have been known for years, and yet organisation after organisation repeats the same script.

In December 2021 Gartner predicted that more than 75 percent of master data management programmes would fail to meet business expectations through 2025 (Magic Quadrant for MDM Solutions). The same figure circulates in the industry with an earlier date and no source attached; this report uses only the documented version.

The causes rarely lie in technology. Across projects and pre-sales conversations we keep meeting four patterns:

1. Scope too wide at the start

The programme covers every data domain from day one. A committee forms, the committee appoints working groups, and after a year the organisation owns workshop documentation instead of a working data layer and a first measurable result.

2. A committee instead of an owner

Nobody holds the mandate to settle a disputed record on its merits. Whether two entries describe the same entity is a business question; without a data owner on the business side, every decision waits for the next meeting.

3. Fully automatic merging

Record matching is probabilistic. Merging two genuinely distinct entities by mistake is harder to detect and costlier to unwind than a duplicate, which is why a person should approve every merge.

4. Cleansing with no rules at the gate

A one-off clean-up without validation rules for new records buys a quarter of order. Then the file drifts back to its previous state, and the organisation concludes that 'we tried this once and it didn't work'.

Four approaches, one condition

The four approaches found on the market differ not in quality but in the scope of responsibility they take on. A one-off cleansing fixes the state of the file on migration day; it does not protect against re-contamination. Warehouses and lakehouses are responsible for collecting and analysing data, not for ruling which record is correct. SAP Master Data Governance is embedded in the SAP ecosystem; in multi-system landscapes it is a separate decision whether an ERP-independent layer is also needed. An ERP-independent master data platform maintains the golden record above the systems, while operational processes rightly stay where they are. These approaches complement one another; trouble starts when a project plan expects one of them to replace the rest.

75%

of master data management programmes will fail to meet business expectations through 2025

Gartner, Magic Quadrant for MDM Solutions, December 2021

8

The incremental approach: method and metrics

The answer to the failure statistics is not a bigger programme but a smaller one: a single data domain, a single success criterion, and results measured on the organisation's own data in weeks, not quarters.

8.1. The order of work

1. Choose the domain and the success criterion

We start with the domain that generates the most manual work and the most arguments about numbers; customers and material masters are the usual candidates. The success criterion is agreed before the start.

2. Measure the current state

On the organisation's real data we establish the number of duplicates, the gaps in critical fields and the discrepancies between systems. Measurement lets the decision rest on numbers rather than opinions.

3. Define the golden-record rules

Name and legal-form normalisation, conflict resolution between sources, versioning, and a data owner with the mandate to decide.

4. Merge under supervision and measure again

Language models are good at suggesting that two entries may describe the same entity, and at explaining why; the decision is approved by a data steward, and every change lands in the audit trail. After the pilot, results are compared against the baseline measurement.

A single-domain pilot usually takes four to eight weeks, depending on the state of the data and the number of systems to feed. This is how we work with Medcover, among others, in the employee data domain: 18 countries, around 45,000 employees and roughly 100,000 data events a year, with validation, an approval workflow and a full history of changes. Typical pilot metrics include the share of duplicates in the domain, the completeness of critical fields, the number of cross-system discrepancies and the time needed to resolve a disputed record.

8.2. A measurement you can run yourselves

The scale of the problem in one domain can be estimated internally within days, before any project decision is made. An extract from the customer master file and four questions are enough:

- 1 How many records share a tax ID but differ in name or address?
- 2 How many records have gaps in fields your processes treat as critical?
- 3 How many entries differ between the ERP and the CRM for the same entities?
- 4 Who decides which version is correct, and how long does that take?

An answer of "we don't know" to any of these questions is also a measurement result - arguably the most instructive one.

8.3. Who this work is for right now

Priority belongs to organisations preparing an S/4HANA conversion, consolidating companies after acquisitions, deploying automation or AI on operational data, or needing to demonstrate to an auditor where their data resides and who has access to it. We advise against starting where the organisation expects a one-off cleansing with no business-side data owner, plans to cover every domain at once before the first confirmed result, or has no one able to settle a disputed record on its merits. Whatever the scope, the programme carries its boundary conditions: access control, personal data protection and retention rules - particularly for employee and customer data. An honest qualification saves both sides a quarter of meetings that lead nowhere.

"In master data pilots the technology is rarely the surprise. What surprises people is how quickly decisions get made once a measurement replaces opinions on the table"

Michał Korzeń, CTO, SNOK

9

Conclusions

Three processes have converged. The market has priced master data management through multi-billion-dollar acquisitions; the Polish regulator has switched on a system that tests every taxpayer's customer master file daily; and vendor support calendars have formed a queue of migrations in which data ranks among the most underestimated risks. In parallel, national research shows that data readiness, not access to technology, is the bottleneck of AI adoption in Polish companies.

The practical conclusion is less glamorous than the covers of AI reports: the advantage will belong not to whoever buys the tools first, but to whoever first puts in order the layer those tools work on. That work can start with a measurement that takes days and requires no investment decision.

10

Methodology and sources

This report was prepared by the SNOK team on the basis of public sources and the firm's own project experience. Market figures were verified in July and August 2026; each is quoted with its source and date. Where sources differ in methodology (GUS versus Eurostat) or originate from a single vendor (invoice error statistics), this is stated explicitly. References to SNOK projects concern completed work; client names appear only with the client's consent. The report is for information purposes and does not constitute tax or legal advice; decisions concerning KSeF and settlements should be reviewed with a tax adviser. Statistical sources are Polish-language publications; titles are given in the original.

- Statistics Poland (GUS), "Społeczeństwo informacyjne w Polsce w 2025 r." (Information Society in Poland 2025), October 2025
- Eurostat, AI use and data analytics in EU enterprises, December 2025 and February 2026
- European Commission, "Poland - 2025 Digital Decade Country Report", June 2025
- EY Poland, "Jak polskie firmy wdrażają AI" (How Polish companies adopt AI; CubeResearch, c. 500 firms, Q4 2025), published 2026
- Algolytics and SW Research, data and AI maturity of Polish companies (700+ firms), 2025
- Ministry of Finance, ksef.podatki.gov.pl: mandatory KSeF timeline and the FA(3) schema guide, 2026
- Director of National Revenue Information (KIS), individual tax ruling ref. 0114-KDIP1-3.4012.44.2026.1.AMA, 24 March 2026 (via jpk.info.pl commentary)
- Ministry of Finance figures on KSeF e-invoices and issuers, via the Polish Press Agency (PAP), May 2026
- Prawo.pl on the second stage of KSeF, 2 April 2026
- Bankier.pl, citing SmartKSeF data by Exorigo-Upos (single-vendor data), 29 July 2026
- SAP, Business Suite 7 support strategy (mainstream to end-2027, extended to 2030), February 2020; transition option 2031-2033, August 2025
- SAP Note 2265093 and official SAP materials on Business Partner and Customer Vendor Integration
- DSAG, Investmentsreport 2026 (n=198), February 2026
- Gartner on ECC customers, via The Register, February 2025
- ISG, "The State of SAP Migration" (n=200), January 2026
- All for One Poland, estimate of SAP ERP users in Poland (company material, undated)
- Salesforce, announcement of the closed Informatica acquisition, 18 November 2025
- SAP, announcements of the Reltio acquisition, 27 March and 7 May 2026
- Gartner, Magic Quadrant for MDM Solutions: December 2021 (MDM programme forecast) and the reinstated April 2026 edition
- Mordor Intelligence (2025) and Grand View Research (2023), MDM market size estimates
- Gartner, press release on AI projects and data management practices, 26 February 2025
- Fordata and Navigator Capital, "M&A Index Poland": 2024 and 2025 summaries
- Microsoft, announcement of the end of support for Dynamics GP, 25 September 2024
- IDC, "Poland Enterprise Application Software Market" (ERP market shares in Poland), 2024

About SNOK

SNOK Sp. z o.o. is a Polish IT consulting firm based in Warsaw. We deliver projects across SAP technology, cybersecurity, automation, and custom software and AI development. Our management systems are certified against ISO 27001:2022 and ISO 9001:2015. We develop SNOK MDM, our own master data management platform in production since 2022, and we advise vendor-neutrally in SAP landscapes.

How an engagement begins

Every master data engagement at SNOK begins with a duplicate measurement: on a sample from a single domain we establish how many records describe the same entity, how many have gaps in critical fields, and where the systems diverge. You receive the results, together with a scope recommendation, within five working days. It is the fastest way to start the data conversation in your organisation from numbers rather than opinions.

The duplicate measurement form and a description of the method are available at snok.ai/en/offer/master-data-management. Questions are welcome directly to the report's authors.

Duplicate measurement · results and a scope recommendation within five working days

Contact the report's authors:

Jacek Bugajski, CEO · jacek.bugajski@snok.ai

Michał Korzeń, CTO · michal.korzen@snok.ai