PRODUCT PRESENTATION · OCTOBER 2026

BeryxTech

Data quality, GDPR and the cost of data in one layer: read at the source, read-only, with the calculation behind every number in plain sight.

For the data team, for the DPO and for whoever pays the cloud bill.

EVERY CLAIM CARRIES A LABEL

● EXISTS TODAY
in production and covered by tests

◐ PARTIAL
exists, with the limits stated on the slide

◐ TO BE SWITCHED ON
built in the code, not yet switched on in production

○ VISION
does not exist yet

Facts checked on 12/10/2026 against the product in production and the sources cited.

Decisions and AI already run on data. Trust in that data has not kept up.

67%

do not fully trust the data they use to make decisions (55% in 2023)

Precisely and Drexel LeBow, Sep 2024 · 565+ professionals · vendor study

64%

name data quality as their top data integrity challenge (50% in 2023)

Precisely and Drexel LeBow, Sep 2024 · vendor study

71%

worry that incorrect or hallucinated outputs reach decision-makers

dbt Labs, State of Analytics Engineering, Apr 2026 · 363 professionals · vendor study

60%

of AI projects without AI-ready data will be abandoned through 2026, according to Gartner's forecast

Gartner, press release of 26/02/2025

When AI answers from unchecked data, the error reaches decision-makers faster.

Three costs that appear in no budget

01 DATA ENGINEERING

The business finds the failure before the data team does

Stale, duplicated or invalid data breaks no job. It shows up at month-end close, in a board report or in an invoice that never goes out.

02 GDPR AND AI

Accuracy of personal data is a legal obligation

The GDPR requires accurate, up-to-date data, answers to data subjects within a month and breach notification within 72 hours. Infringing the principles carries fines of up to €20M or 4% of worldwide annual turnover. The AI Act adds requirements on training data from December 2027.

GDPR, Arts. 5, 12, 33 and 83; AI Act, Arts. 10 and 113

03 DATA FINOPS

The data bill grows with no owner

Tables nobody reads keep being loaded. One badly written query multiplies the day's cost and only shows up in the monthly bill. You pay compute to process data that fails its checks.

The three costs share one root: nobody measures the data continuously, with an owner and with evidence.

Data fails silently: the jobs stay green and the numbers go wrong

ANATOMY OF A DATA INCIDENT

ILLUSTRATIVE EXAMPLE · SYNTHETIC DATA

FRIDAY, 6 PM

The marketplace connector stops syncing. No job fails: rows simply stop arriving.

WEEKEND

Marketplace orders never reach the data warehouse. The sales dashboard shows a drop that never happened.

MONDAY, 10 AM

The weekly close does not add up. The investigation starts: which table, since when, and who else uses this data?

TUESDAY

With the source found, it remains to work out which reports and models have already consumed the wrong data.

SIX WAYS TO FAIL WITHOUT AN ERROR

Freshness
the table stopped receiving data

Volume
half the rows arrived

Schema
a column changed type or name

Duplicates
a retry wrote the same order twice

Validity
an 8-digit tax number, an IBAN with the wrong check digit

Integrity
orders without a customer after a delete without cascade

The story and the examples come from BeryX's demo packs, with synthetic data. Each failure maps to a monitor or a rule in the product.

The cost of not knowing is counted with your numbers

PILLAR	WHAT TO COUNT	HOW TO ESTIMATE	WHAT BERYX MEASURES
Data engineering	Data incidents a month, hours of investigation and days to detection	Incidents × hours × hourly cost, plus days to detection × value at risk per day (invoices not yet issued, for example)	History of issues and incidents with dates, and estimated impact in euros per issue, labelled as a heuristic
GDPR	Data subject requests a month and datasets with personal data but no owner or no rules	Requests × hours per request; maximum exposure of €20M or 4% of worldwide turnover	Personal data map with the rule coverage of each column
Data FinOps	Tables unread for 90 days, compute spent on data that fails its checks, and daily spikes	Monthly cost of those tables, plus the compute attributed to datasets with open issues	Cost read from the platform's system views, per day (per dataset on BigQuery, per account on Snowflake and Databricks), cost spikes and real usage per table

Sources: GDPR, Art. 83(5)

In a first conversation we fill in this table with you, using the data from your own sources.

What the law asks of data, and that can only be proven with measurements

OBLIGATION	WHAT IT REQUIRES	DEADLINE OR LIMIT	WHAT IT DEMANDS OF THE DATA
Accuracy · GDPR, Art. 5(1)(d)	Personal data accurate and kept up to date; inaccurate data erased or rectified without delay	Fine of up to €20M or 4% of worldwide turnover (Art. 83(5))	Knowing which personal data is wrong, where and since when
Data subject rights · Art. 12(3)	Answer requests for access, rectification or erasure	One month, extendable by two more	Knowing which systems hold each person's data
Personal data breach · Art. 33(1)	Notify the supervisory authority (in Portugal, the CNPD)	72 hours, where feasible	Knowing which data and which data subjects were affected
Impact assessment · Art. 35	Assess before high-risk processing	Before the processing starts	An inventory of data categories and data flows
Data for high-risk AI · AI Act, Art. 10(3)	Training, validation and testing data that is relevant, representative and, to the best extent possible, free of errors and complete	From 2 December 2027 (Annex III)	Measuring errors and completeness in the data sets

A summary for context; not legal advice.

Sources: GDPR, Regulation (EU) 2016/679 · AI Act, Arts. 10 and 113 (AI Act Service Desk)

Portugal's data protection authority has already fined bodies millions over personal data handling

€4.3M

Statistics Portugal (INE)

DECEMBER 2022

2021 Census. Five infringements: health and religion data processed without a clear lawful basis, insufficient information to respondents, a US processor chosen without due diligence, international transfers without supplementary measures and no impact assessment.

STATUS INE announced it would challenge the decision in court.

CNPD, Deliberation/2022/1072 · Observador, 13/12/2022

€1.25M

Lisbon City Council

JANUARY 2022

Data of protest organisers sent to embassies and to the Ministry of Foreign Affairs: 225 infringements.

STATUS The courts upheld the infringement. With part of the infringements time-barred, the amount fell to €738k at the Central Administrative Court South (August 2025), in a decision that can still be appealed.

Renascença, 14/01/2022 · Observador, 06/08/2025

€400k

Barreiro-Montijo Hospital Centre

2018

The first GDPR fine in Portugal: 985 accounts with a doctor profile for 296 doctors, and improper access to clinical records.

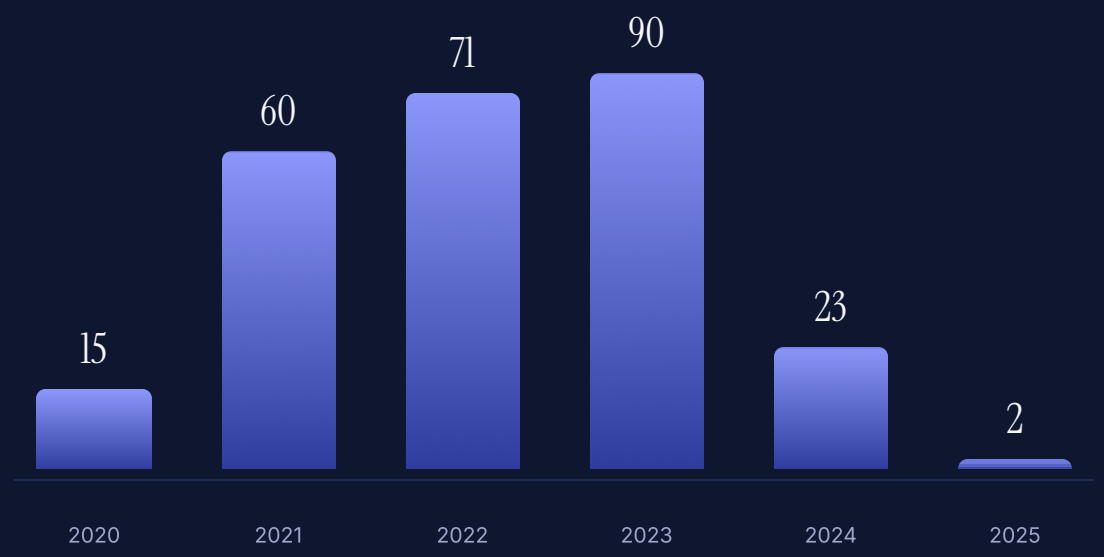
STATUS It became a single fine of €380k. In July 2020, the CNPD agreed to waive it, citing the pandemic and the hospital's finances.

Garrigues, 2018 · Renascença, 27/05/2021

We do not claim BeryX would have prevented these fines. The three cases show what the CNPD investigates: where sensitive data lives, who accesses it and where it goes.

Fines rise and fall with the CNPD's resources. Notified breaches keep rising.

FINES ISSUED BY THE CNPD, PER YEAR



472

personal data breach cases in 2025, 42% more than in 2024 (357 in private bodies, 115 in public ones)

CNPD, 2025 Activity Report, pp. 15 and 34

88

infringement proceedings opened in 2025 which, in the CNPD's words, could have led to 90 fines instead of 2

CNPD, 2025 Activity Report, p. 42

2,037

inquiries opened in 2025

CNPD, 2025 Activity Report, p. 42

> The CNPD itself attributes the 2025 drop to a shortage of staff to handle cases and, in September 2025, proposed a faster infringement regime to Parliament. The obligations stay the same: notify within 72 hours, answer within a month, prove accuracy.

Sources: CNPD, 2025 Activity Report, approved on 24/03/2026 (fig. 10, pp. 12, 42 and 43)

The data bill grows with no owner and no link to quality

84%

say managing cloud spend is the top cloud challenge for organisations today

Flexera, State of the Cloud 2025 · 750+ professionals · vendor study

- × Tables nobody reads keep being loaded, stored and monitored.
- × One badly written query multiplies the day's cost, and it only shows up in the monthly bill.
- × You pay compute to process data that fails its checks.
- × AI costs grow with no cap and no split by team.

The bill lands with finance and the quality issues stay with the data team. BeryX puts the two side by side, per dataset wherever the platform allows it.

Where does the truth about your data live today?

Today

- × In tests and scripts scattered across teams and repositories
- × In the memory of whoever knows the tables
- × In a spreadsheet record of processing, reviewed once a year
- × In the cloud bill, at the end of the month
- × In dashboards nobody links back to where the data comes from

With BeryX

- ✓ In versioned rules and contracts that run on every scan • EXISTS TODAY
- ✓ In a catalogue with owner, domain, sensitivity and usage per dataset • EXISTS TODAY
- ✓ In a personal data map updated on every scan • EXISTS TODAY
- ✓ In the daily cost read from the system views of BigQuery, Snowflake and Databricks • EXISTS TODAY
- ✓ In incidents with lineage, an owner and a deadline • EXISTS TODAY

One trust record per dataset, shared by engineering, the DPO and finance.

Why now

2025

Personal data breaches notified to the CNPD rise by 42%, to 472 cases.

CNPD, 2025 Activity Report

SEP 2025

The CNPD proposes a faster infringement regime to Parliament, to stop cases becoming time-barred.

CNPD, 2025 Activity Report, p. 12

2026

Gartner forecast: organisations abandon 60% of AI projects that lack AI-ready data.

Gartner, 26/02/2025

2 AUG 2026

The AI Act applies in general (Art. 113).

AI Act Service Desk, Art. 113

2 DEC 2027

High-risk obligations for Annex III, including the quality of training, validation and testing data (Art. 10).

AI Act Service Desk, Arts. 10 and 113

2 AUG 2028

High-risk obligations for Annex I, the regulated products.

AI Act Service Desk, Art. 113

> Getting data into shape takes months: profiling, rules, owners and a history of measurements that shows progress.

A trust layer over the data you already have, without copying it

YOUR SOURCES STAY WHERE THEY ARE

- Databases
- Data warehouses
- Files and object storage
- Business applications: SAP, Salesforce, Dynamics 365
- Documents
- Streaming

→

BERYX · READ-ONLY

Discover | Validate | Monitor | Explain | Prove

- ✓ Aggregates computed at the source
- ✓ Limited samples, with retention
- ✓ Proposes fixes, never writes to the sources
- ✓ Every number says how it was calculated

→

Data engineering
incidents with lineage, SLOs, rules as code

DPO and legal
personal data map, sensitivity, evidence

FinOps and leadership
cost and quality side by side, real usage, reports per role

WHERE THE TEAM ALREADY WORKS: Slack | Teams | Jira | Power BI | Email | Webhooks | API | MCP

30 source types in the connector registry; 1 not yet installed in the SaaS, Microsoft Fabric (see Ecosystem).

From connected data to proven data, and back to the start

01
Connect
read-only credentials

→

02
Discover
profile, personal data, usage

→

03
Rules
proposed by the profile, accepted by people

→

04
Monitor
freshness, volume, schema, anomalies, cost

→

05
Incident
lineage, impact, owner, deadline

→

06
Fix
SQL or a pull request, rehearsed first

→

07
Prove
reports, evidence, audit

What is learned flows back into the rules: thresholds, flagged false positives and rules suggested from resolved issues.

Every step is recorded: who decided, when and on what evidence.

Every dataset carries its own trust record

Owner and domain
who answers for it and which area it belongs to

EXISTS TODAY

Sensitivity per column
public, internal, confidential or personal data, with author and date

EXISTS TODAY

Rules and contract
active rules and a versioned ODCS contract

EXISTS TODAY

Score and SLO
five dimensions, trend and service objective

EXISTS TODAY

Real usage
queries per day: hot, active, cold or unused

EXISTS TODAY

Cost
attributed compute, when the platform reports it

EXISTS TODAY

Incidents and lineage
what failed, upstream and downstream

EXISTS TODAY

Audit
every change, written in the same transaction

EXISTS TODAY

> The engineer, the DPO and the finance director read the same record. Nobody has to explain the context again.

Connect

INPUT

A read-only credential for a database, a warehouse, files or an application.

WHAT BERYX DOES

Tests the connection, discovers schemas and tables and imports the metadata. The credential is encrypted before it is stored and the address is checked against SSRF before the first request.

● EXISTS TODAY

WHO DECIDES

The workspace administrator chooses the source, the scope and how often scans run.

GOVERNANCE AND EVIDENCE

Read-only enforced in code: the HTTP client refuses writes before they leave, and Postgres sessions open in read-only mode. In tests, real SQL Server, Oracle and MongoDB servers refused writes.

● EXISTS TODAY

OUTPUT

A catalogue of datasets with schema, schedule and encrypted credential.

● EXISTS TODAY

CONNECTS TO

PostgreSQL · SQL Server · Oracle · BigQuery · Snowflake · Databricks · S3 · SAP · Salesforce · Dynamics 365

HANDED TO THE NEXT STEP

The list of datasets and columns goes on to profiling.

VALUE

No write credentials, no software installed on the sources and no copied tables.

● EXISTS TODAY

Discover

INPUT

The connected datasets.

WHAT BERYX DOES

Column profiling with aggregates computed at the source: nulls, distinct values, patterns and distributions. Recognises tax numbers, IBANs, emails and postcodes, builds the personal data map, classifies sensitivity on four levels and estimates re-identification risk (k-anonymity). On platforms with telemetry, it measures the real usage of each table.

● EXISTS TODAY

WHO DECIDES

The DPO confirms or corrects the sensitivity of each column. A scan never overrides a manual decision.

GOVERNANCE AND EVIDENCE

Every number says whether it is exact, estimated or sampled, with a 95% confidence interval. Every classification change is recorded with author and date.

● EXISTS TODAY

BeryX methods

OUTPUT

Profile, score per dimension, personal data map, sensitivity per column and usage.

● EXISTS TODAY

CONNECTS TO

Semantic types · PII map · k-anonymity · BigQuery · Snowflake

HANDED TO THE NEXT STEP

The profile feeds the suggested rules.

VALUE

The DPO sees where personal data lives, read from the sources themselves, without relying on a manual inventory.

● EXISTS TODAY

Rules and contracts

INPUT

The profile, the suggestions and the tests the team already has.

WHAT BERYX DOES

Sixteen rule types: formats with check digits (tax numbers, IBANs, cards), uniqueness, ranges, reference lists, referential integrity, freshness, volume and custom read-only SQL. Accuracy against external sources, opt-in: VIES, GeoAPI.pt, numbering plans and MX records. Imports dbt tests, Great Expectations, SodaCL and ODCS contracts.

● EXISTS TODAY

WHO DECIDES

The profile and AI propose; a person accepts each rule.

GOVERNANCE AND EVIDENCE

Customer SQL goes through a guard: a single SELECT, in a transaction that is always rolled back. Configuration also lives as code, in a versioned datavault.yaml.

● EXISTS TODAY

OUTPUT

Active rules per dataset and a versioned contract; export to dbt, with SodaCL and Great Expectations still to be switched on.

◐ PARTIAL

CONNECTS TO

dbt · Great Expectations · SodaCL · ODCS · VIES · GeoAPI.pt

HANDED TO THE NEXT STEP

Accepted rules run on every scan and feed the monitors.

VALUE

Rules stop living in one person's scripts.

● EXISTS TODAY

Monitor

INPUT

Rules, profile history and cost and usage telemetry.

WHAT BERYX DOES

Automatic freshness, volume and schema monitors, up to 500 per connection on every scan. Anomalies with a modified z-score and false-discovery control (Benjamini-Hochberg), distribution drift, an SLO per dataset and daily cost spikes, with the day's most expensive queries on BigQuery.

● EXISTS TODAY

WHO DECIDES

The dataset owner sets the SLO and the severity. A false positive marked «ignored» silences the series.

GOVERNANCE AND EVIDENCE

Published methods, tested against the code. No quality, anomaly or cost figure comes from an LLM.

● EXISTS TODAY

OUTPUT

Alerts in the app, by email, in Slack, in Teams and by webhook, with on-call escalation.

● EXISTS TODAY

CONNECTS TO

Slack · Teams · Email · Signed webhooks

HANDED TO THE NEXT STEP

An alert opens or updates an issue, which can join an incident.

VALUE

The failure reaches the data team first, not the month-end close.

● EXISTS TODAY

Incidents and root cause

INPUT

An alert or an open issue.

WHAT BERYX DOES

Groups issues into incidents through lineage (foreign keys, view SQL, dbt manifest, OpenLineage) and shows the downstream impact. AI proposes the likely cause from metadata and limited samples. Response deadlines by severity, with escalation.

● EXISTS TODAY

WHO DECIDES

The domain owner accepts, assigns and closes the incident.

GOVERNANCE AND EVIDENCE

Audit trail written in the same transaction as the change and exportable to the SIEM: Splunk, Sentinel or Google SecOps.

● EXISTS TODAY

OUTPUT

An incident with lineage, impact, owner, deadline and history, and a ticket in Jira, Linear, ServiceNow, Azure DevOps, GitHub or GitLab.

● EXISTS TODAY

CONNECTS TO

Jira · Linear · ServiceNow · Azure DevOps · GitHub · GitLab · OpenLineage

HANDED TO THE NEXT STEP

The likely cause and the impact go on to the proposed fix.

VALUE

Knowing which reports and models were affected before anyone asks.

● EXISTS TODAY

Fix

INPUT

An incident with a likely cause.

WHAT BERYX DOES

Proposes the fix in SQL, in the source's dialect, or in PySpark, and rehearses it on a sample before it is accepted. From the Business plan, the approved fix opens a pull request in the customer's GitHub or GitLab. Standardisation and deduplication also come out as proposals.

● EXISTS TODAY

WHO DECIDES

A person approves and the customer applies it through their own CI. BeryX never writes to the sources.

GOVERNANCE AND EVIDENCE

A proposed, accepted, applied and verified flow: the next scan confirms whether the issue is gone.

● EXISTS TODAY

OUTPUT

A pull request or approved SQL, linked to the dataset and the ticket.

● EXISTS TODAY

CONNECTS TO

GitHub · GitLab · SQL · PySpark · dbt

HANDED TO THE NEXT STEP

The next scan's verification closes the loop and goes into the history.

VALUE

The fix is reviewed and versioned like any other code.

● EXISTS TODAY

Prove

INPUT

The history of scans, incidents and decisions.

WHAT BERYX DOES

Reports per role (engineering, management, leadership) with estimated impact in euros, as PDF and on a schedule; a CDO panel with targets and traffic lights; the AI Act Art. 10 dossier (Business); feeds for Power BI; audit export; a public badge per data product.

● EXISTS TODAY

WHO DECIDES

Leadership reads; the DPO and internal audit use the evidence.

GOVERNANCE AND EVIDENCE

The impact in euros is a declared heuristic that tends to underestimate. Every number says how it was calculated.

● EXISTS TODAY

OUTPUT

Reports, dossiers and exports with date and source.

● EXISTS TODAY

CONNECTS TO

Power BI · Tableau · Looker · PDF · CSV

HANDED TO THE NEXT STEP

What is learned flows back to the rules and thresholds of step 3.

VALUE

Answering an auditor or the regulator with history, not with a hurried reconstruction.

● EXISTS TODAY

Monday at the online shop: from a stalled connector to the weekly report

A demonstration with BeryX's «Loja online» pack: synthetic data from a Portuguese online shop, with eight planted issues.

ILLUSTRATIVE EXAMPLE · SYNTHETIC DATA

01

Connect

The orders database connects read-only. The first scan profiles customers, products, orders and order lines in seconds.

02

Discover

The profile flags seven columns with personal data (name, email, tax number and address) and the format rule reports 4% invalid emails.

03

Rules

The e-commerce pack installs rules for email, ISO country, postcode, order uniqueness and freshness. The profile proposes more rules, which a person accepts or rejects.

04

Monitor

The freshness rule fails: the most recent order is 72 hours old, against a limit of 24. The marketplace connector stopped on Friday.

05

Incident

Lineage groups related issues into incidents, each with its origin, severity and the datasets affected.

06

Fix

For the 14 repeated order_id values, BeryX proposes the diagnostic SQL and the fix, also in PySpark. The fix is flagged to run only after the diagnosis is validated.

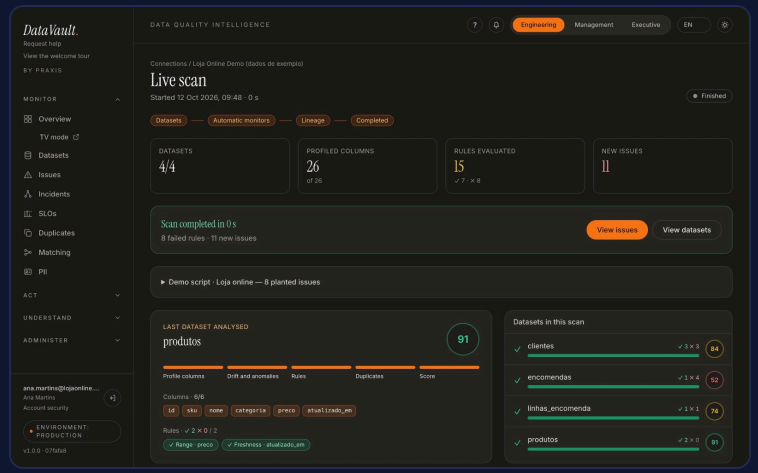
07

Prove

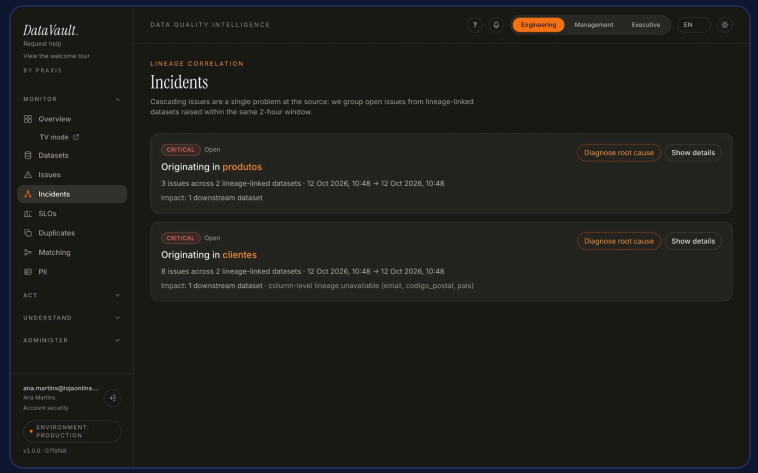
The executive report sums up the business risk, the estimated financial exposure, the trend and the critical incidents, with the recommended action.

Reproducible: the free account includes the one-click demo with this pack and three more (ERP, Neobanco and Hospital).

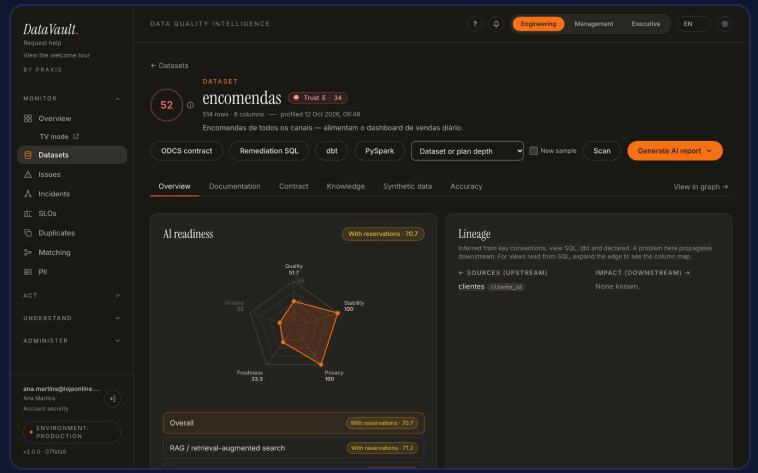
What the team sees



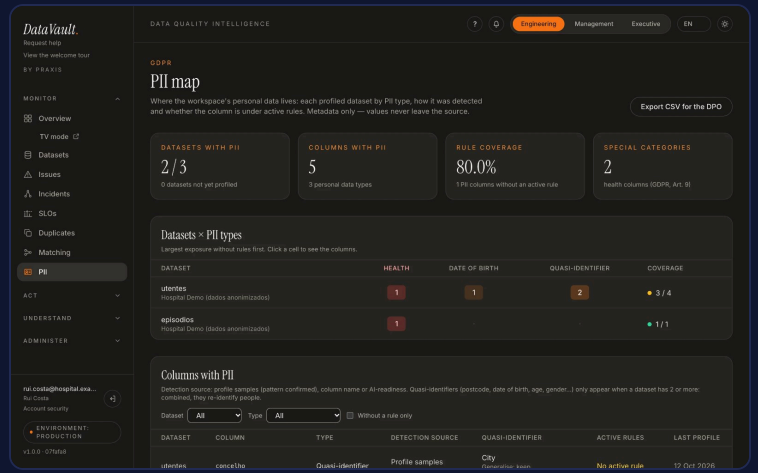
Live scan: failed rules, new issues and a score per dataset



Incidents: issues grouped through lineage, with the origin and the impact



Dataset detail: score per dimension, AI readiness and lineage



Workspace personal data map, with special categories (health)

Screenshots of the running app with the demo packs (synthetic data).

Reference architecture

PEOPLE AND TOOLS

Web app | Slack and Teams | Power BI, Tableau and Looker | AI agents over MCP | Versioned public API

BERYX · GOOGLE CLOUD, EUROPE-WEST1

API | Asynchronous workers for scans and reports, with queues and a DLQ

Postgres holding the metadata, never the tables | AI facade with a per-task router

MCP server with 21 read-only tools

YOUR SOURCES

Read-only connections through each vendor's official methods

> What leaves the sources: aggregates, up to 20 example values per column and, for row rules, a sample of up to 5,000 rows per table.

- Shared SaaS in the EU (europe-west1) — EXISTS TODAY
- Dedicated instance per customer, with its own egress IP and optional CMEK — TO BE SWITCHED ON
- Running inside the customer's perimeter — VISION

Subprocessors and regions are published on the security page, with 30 days' notice before any change.

AI with rules: where it comes in, where it does not, and who decides

Where AI comes in

- Rule suggestions, likely cause of incidents and reports per role — EXISTS TODAY
- A natural-language assistant over the catalogue and the issues — EXISTS TODAY
- Glossary and dataset documentation, always for human approval — EXISTS TODAY

Where it does not

- No quality, anomaly or cost figure comes from an LLM: the methods are deterministic and published — EXISTS TODAY
- Proposing never executes: every action asks for human approval and is audited — EXISTS TODAY

Customer controls

- AI switched off on any plan, with reports in offline mode — EXISTS TODAY
- Raw or synthetic samples, as the workspace chooses — EXISTS TODAY
- Your own Anthropic key (Business) or Claude on Amazon Bedrock or Vertex AI, in European regions (Enterprise) — EXISTS TODAY

Guards

- Daily AI cost cap and a monthly token quota per plan — EXISTS TODAY
- Per-task router: by default, only Anthropic receives customer data — EXISTS TODAY
- Output guard that masks secrets, emails, IBANs, cards and tax numbers before storing (it does not yet cover the streaming assistant) — PARTIAL
- Defence against indirect prompt injection, with adversarial tests in CI — EXISTS TODAY
- Marking of AI-generated content, as AI Act Art. 50 requires — EXISTS TODAY

TO BE SWITCHED ON — By default, limited samples go to Anthropic, in the US, under standard contractual clauses. The anonymiser (Presidio) is built but not yet in the production image.

GDPR: from a manual inventory to a map that updates on every scan

Knowing where personal data lives

- Personal data map: which datasets hold which types of data, how each was detected and whether it has active rules — EXISTS TODAY
- Sensitivity per column on four levels, with inheritance through lineage (suggested) and curation with author and date — EXISTS TODAY
- Re-identification risk from quasi-identifiers (k-anonymity); examples sent to AI are generalised — EXISTS TODAY

Keeping it accurate

- Check digits for tax numbers, IBANs and cards, and reference lists — EXISTS TODAY
- Accuracy against VIES, GeoAPI.pt, numbering plans and MX records, opt-in — EXISTS TODAY
- Accuracy audit by human sampling, with a confidence interval — EXISTS TODAY

Answering and proving

- A register of data subject requests with each jurisdiction's legal deadline: EU, United Kingdom, Switzerland, Brazil, California and Angola — TO BE SWITCHED ON
- Searching for and erasing a data subject's value in the samples BeryX stores — TO BE SWITCHED ON
- Sample retention configurable per workspace and an audit trail exportable to the SIEM — EXISTS TODAY
- Scanning for forgotten credentials in code and notebooks — EXISTS TODAY

The CNPD cases, as questions the data answers

Where is the health, religious or financial data?

The personal data map and sensitivity per column. EXISTS TODAY

Who has access, and does it match who should?

A referential integrity rule between the account directory and the staff register. EXISTS TODAY

Where does the data go?

Lineage between the connected sources. Data sent by email or to third parties stays out of scope. PARTIAL

Data FinOps: cost read from the platform itself and crossed with quality

Cost and quality side by side (Business)

- Native telemetry: BigQuery (INFORMATION_SCHEMA.JOBS), Snowflake (WAREHOUSE_METERING_HISTORY) and Databricks (system.billing.usage) — EXISTS TODAY
- How much of the month's compute ran on data that fails its checks — EXISTS TODAY
- Cost per table on BigQuery; on Snowflake and Databricks the cost belongs to the account, and the split is labelled as an attribution. Without telemetry, no number appears at all — EXISTS TODAY

Cost spikes

- A daily spike opens an issue and alerts the channels; on BigQuery, with the day's 5 most expensive queries and their literals masked — EXISTS TODAY
- Published criterion: z ≥ 3, at least 50% and 1 USD above the median, with 7 days of history — EXISTS TODAY

Real usage

- Hot tables (20 or more queries a day), active, cold and unused (0 queries in 90 days, only with telemetry covering at least 80 days) — EXISTS TODAY
- Unused tables become candidates for decommissioning — EXISTS TODAY
- Usage-driven scans: hot tables deeply on every scan, unused ones only in the monthly profile, opt-in and approved by the steward — EXISTS TODAY

BeryX's own cost

- A bytes cap before reading on BigQuery — EXISTS TODAY
- AI cost per report, per task and per provider; scheduled reports through the batch API at half price — EXISTS TODAY
- Chargeback of usage per workspace, domain and team, with a monthly CSV and a Power BI feed — EXISTS TODAY

PARTIAL — Cost telemetry is built and tested, but not yet proven on a real BigQuery, Snowflake or Databricks account. It measures compute cost, not how efficient the data format is.

Sources: BeryX methods, cost anomalies

Security: read-only enforced in code, and the gaps put in writing

Access to the sources

- Read-only in the code, tested against real servers — EXISTS TODAY
- Each vendor's official methods, and credentials encrypted (Fernet) with key rotation — EXISTS TODAY
- Anti-SSRF with pinned DNS, for webhooks and tickets too — EXISTS TODAY

Isolation and identity

- A test walks every API route looking for leaks between customers — EXISTS TODAY
- SSO over OIDC and SAML 2.0 on every plan; MFA mandatory for organisation administrators — EXISTS TODAY
- An IP allowlist per organisation, for API tokens too — EXISTS TODAY

Trail and control

- Audit written in the same transaction as the change, with export to the SIEM — EXISTS TODAY
- BeryxTech support access only with consent, read-only and for 72 hours at most — EXISTS TODAY
- Export of the workspace's data and cascading deletion with a signed certificate — EXISTS TODAY

What we do not have yet

- × We do not have ISO 27001 or SOC 2 yet: both are in progress, with no dates.
- × We have not had an external penetration test yet.
- × WAF and Cloud Armor are written but not yet active in production.
- × No contractual availability or support SLA yet.
- × The contract templates (DPA and MSA) have not yet been reviewed by a lawyer.

Ecosystem: 30 source types and the tools the team already works in

DATABASES 7

PostgreSQL · MySQL and MariaDB · SQL Server, Azure SQL and Synapse · Oracle · Amazon Redshift · MongoDB · SQLite

WAREHOUSES 4

BigQuery · Snowflake · Databricks · Microsoft Fabric *

FILES AND OBJECT STORAGE 6

CSV · Google Sheets · Google Cloud Storage · Amazon S3 · Azure Blob and ADLS Gen2 · SFTP

BUSINESS APPLICATIONS 5

Salesforce · SAP S/4HANA and SAP Gateway · OData v2 and v4 · Dynamics 365 · REST and JSON

DOCUMENTS 4

Confluence · Jira · SharePoint · Google Drive

STREAMING 2

Kafka · Pub/Sub

COLLABORATION 2

Slack · Teams

* Code done; the driver is not yet in the SaaS image.

- **Catalogues** · OpenMetadata and DataHub; Collibra and Purview validated only against a simulated transport — PARTIAL
- **BI** · Power BI, Tableau and Looker — EXISTS TODAY
- **Tickets** · Jira, Linear, ServiceNow, Azure DevOps, GitHub and GitLab — EXISTS TODAY
- **Orchestration** · dbt Cloud, dbt Core, Airflow, Fivetran and Airbyte — EXISTS TODAY
- **AI agents** · Published MCP server, with 21 read-only tools — EXISTS TODAY
- **Developer packages** · Python SDK, Terraform provider, VS Code extension, Airflow and Claude Code plugin, not yet published — TO BE SWITCHED ON
- **SIEM** · Splunk, Microsoft Sentinel and Google SecOps — EXISTS TODAY

Proof per connector (matrix of 10/10/2026): 7 against real public sources, 2 against real services in Docker, 13 with a simulated transport, 6 without a driver in the test environment and 1 still to verify.

Every number says how it was calculated

Three kinds of number

exact, estimated or sampled, marked in the app

EXISTS TODAY

Confidence intervals

at 95%, Wilson with finite population correction; Fisher's z for correlations

EXISTS TODAY

Score with fixed weights

completeness 25%, uniqueness 25%, validity 20%, timeliness 15% and consistency 15%

EXISTS TODAY

Anomalies

modified z-score (median and MAD) with Benjamini-Hochberg false-discovery control

EXISTS TODAY

Impact in euros

a declared heuristic, always estimated, that tends to underestimate

EXISTS TODAY

Public methods

the document is published, and a test fails if a documented constant changes in the code

EXISTS TODAY

> No calculation method uses an LLM.

Sources: BeryX methods

Six roles, each with its own guide and its own tour

Workspace administrator
Connects sources, manages the team, the alert channels and the AI policy.

● EXISTS TODAY

Role guide →

Member
Works on issues, rules, suggestions and incidents.

● EXISTS TODAY

Role guide →

Viewer
Reads scores, reports and the catalogue without changing anything. Cannot yet be assigned on the Team page.

◐ PARTIAL

Role guide →

Domain owner
Decides on the datasets, incidents and rules of their domain.

● EXISTS TODAY

Role guide →

Organisation administrator
Governance of every workspace: SSO, IP allowlist, data policy and chargeback.

● EXISTS TODAY

Role guide →

Billing
Plan, invoices and the cost report per team.

● EXISTS TODAY

Role guide →

At first sign-in, a tour per role shows the app. The guides are published in the BeryX docs.

What changes for each function

FUNCTION	TODAY	WITH BERYX
Data engineering	Finds failures through the business; tests scattered across repositories	Automatic monitors, incidents with lineage, versioned rules and a CI gate
DPO and legal	A manual inventory, reviewed once a year	A personal data map on every scan, sensitivity with an author and exportable evidence
FinOps and finance	The bill at the end of the month, with no link to quality	Daily cost per dataset or per account, cost spikes with the most expensive queries, and unused tables
Leadership and CDO	Reports with no source and no trend	A panel with targets and traffic lights, and reports per role
AI teams	Training data nobody measures	Readiness per purpose, the Art. 10 dossier and synthetic data with a privacy assessment

The landscape: where each approach is strongest

Enterprise suites · Collibra, Informatica

- Catalogue and glossary as the company's system of record
- ISO 27001 and SOC 2, with a public trust centre
- Cleansing and MDM that write to your systems, with write credentials

COST Public price on UK G-Cloud: £74,999 to £122,919 a year, plus implementation

UK G-Cloud — Collibra Data Quality (service 136629392504870) · UK G-Cloud — Informatica IDMC (service 727452156234281) · Collibra — Trust Center

Open source · Great Expectations, Soda

- Free and running inside your perimeter
- Unlimited SQL and Python in the tests
- Every test is written and maintained by your team

COST Licence 0; Soda Team at $750 a month, plus compute and engineering hours

Great Expectations — plans and pricing · Soda — plans and pricing

Observability · Monte Carlo, Sifflet

- Mature metadata monitors on cloud warehouses
- SOC 2 Type II and ISO 27001
- They detect and alert; the fix stays with the team

COST Annual contracts; Sifflet at about $48k a year on its published price

Monte Carlo — docs: security and compliance overview · Monte Carlo — Trust Center · Sifflet — plans and pricing

BeryX

- Quality, personal data and cost measured in the same catalogue
- Proposes the fix in SQL or as a pull request, without write credentials
- No certifications or contractual SLA yet

COST Free · from €490/month

Full matrix, with a dated source in every cell →

A summary of the site's comparison matrix, with a dated public source in every cell (checked on 28/09/2026).

Where the difference is, and what we do not claim

What can be tested today

- Quality, personal data and cost measured over the same catalogue and the same history
- Read-only enforced in code, with fixes as proposals and pull requests
- Every number with a published method and a confidence interval
- AI with output rules, an offline mode and the option to switch it off
- A published price and a free account with no sales call

What we do not claim

- ✓ That BeryX would have prevented a particular fine
- ✓ Certifications: we do not have ISO 27001 or SOC 2 yet
- ✓ A contractual SLA
- ✓ That every connector has been proven on a real account
- ✓ That it replaces an MDM or an enterprise catalogue
- ✓ That it fixes the data on its own

> The proof we propose is a pilot, with metrics measured before and after.

What becomes measurable

METRIC	HOW BERYX MEASURES IT	STATUS
Quality score per dataset and per domain, with trend	Five dimensions with fixed weights, on every scan	EXISTS TODAY
SLO compliance per dataset	A service objective per dataset and its compliance history	EXISTS TODAY
Open and resolved issues by severity	Issues and incidents with dates, owner and response deadline	EXISTS TODAY
Columns with personal data but no rules or no owner	The personal data map crossed with rules and owners	EXISTS TODAY
Deadline of each data subject request	A request register with the jurisdiction's deadline	TO BE SWITCHED ON
Compute spent on data that fails its checks	The platform's cost telemetry crossed with the score	EXISTS TODAY
Tables unused for 90 days	Query counts per table from the system views	EXISTS TODAY
AI cost per report and per team	Cost per call, per task and chargeback per team	EXISTS TODAY

Baselines come from the first month of measurement; we do not use market averages.

The honest ledger

● EXISTS TODAY

- Profiling, 16 rule types, automatic monitors, anomalies and SLOs
- Incidents with lineage, likely cause and tickets
- Proposed, rehearsed fixes and pull requests (Business)
- Personal data map, sensitivity and re-identification risk
- Cost and quality side by side, cost spikes, real usage and chargeback
- Reports per role, the assistant and the MCP server
- SSO over OIDC and SAML, MFA, IP allowlist, audit and SIEM
- A tour and a written guide per role

◐ PARTIAL

- Connectors not yet proven on a real account, BigQuery, Snowflake, Databricks and Oracle among them
- Collibra and Purview validated only against a simulated transport
- Metadata pages and the critical data element dossier (BCBS 239) only through the API
- Entity matching on samples of up to 50,000 rows
- The AI output guard does not yet cover the streaming assistant
- The Viewer role cannot yet be assigned on the Team page

○ TO BE SWITCHED ON

- Data subject requests and export to SodaCL and Great Expectations: built, in the next deploy
- Seven connectors not yet in the SaaS image
- The anonymiser in the production image
- Dedicated instance, high availability, WAF and a fixed egress IP
- SDK, Terraform, VS Code and Airflow packages not yet published
- Online payment

○ VISION

- ISO 27001, SOC 2, ISO 27701 and ISO 42001 certifications: we do not have any yet
- External penetration test
- SCIM and security keys (WebAuthn)
- Private connectivity and running inside the customer's perimeter
- Two-way sync with ServiceNow

Plans

Free

€0 /month

- 1 source
- 25 datasets
- 5 AI reports a month
- 30 days of history

Team

€490 /month

- 5 sources
- 500 datasets
- 200 AI reports a month
- 365 days of history
- ✓ API tokens
- ✓ Alert webhooks
- ✓ Document analysis for RAG
- ✓ Scheduled email reports
- ✓ Synthetic data from the profile

Business

€1,190 /month

- 15 sources
- 2,000 datasets
- 1,000 AI reports a month
- 730 days of history
- ✓ AI Act compliance dossier
- ✓ Platform cost telemetry
- ✓ Remediation as a Pull Request (GitHub/GitLab)
- ✓ Your own Anthropic key (BYOK)
- ✓ Choice of AI models per task

Enterprise

On request

- unlimited
- ✓ Claude in your organisation's cloud account (Amazon Bedrock or Google Vertex AI)

> On every plan: the quality engine, suggestions, offline mode, AI that can be switched off, and SSO.

Prices exclude VAT, billed annually; monthly billing costs 20% more. Until online payment is live, subscriptions are arranged with BeryxTech.

How to start

01

Try it on your own

A free account with one source and the one-click demo, with four synthetic data packs: Loja online, ERP, Neobanco and Hospital.

Create a free account →

02

A conversation with BeryxTech.

Thirty minutes to fill in the cost table with your numbers and see the demo on your case.

Talk to us →

03

A pilot on one domain

One data domain, the metrics from the previous slide measured at the start and at the end, and a decision made with numbers.

Propose a pilot →

Data that engineering, the DPO and finance trust, with the same evidence.

BeryX, a BeryxTech product.

beryxtech.com/en/apresentacao