

PILOT DOCUMENTATION

Data Preparation Guide

Part of the ReguShield AI Pilot Success Kit — everything your team needs to run a successful compliance-intelligence pilot.

This guide explains exactly how to prepare your data so ReguShield AI can analyse it cleanly. Getting the file right is the single most important step in the pilot — a few minutes of preparation produces clear, complete results.

1. The Golden Rule: One Flat File, One Row Per Record

ReguShield AI analyses a single flat file per dataset, with one row per record.

The engine **does not join** separate customer and transaction files. Do not upload one file of customers and another of transactions and expect them to be linked — they will not be. Put everything for each record on its own row, in one file.

If you have related information spread across multiple files, flatten it into a single table first, where each row is self-contained.

2. Accepted Formats

FORMAT	STATUS	NOTES
CSV	Primary ("Live") path	Recommended.
Excel (.xlsx / .xls)	Primary ("Live") path	Fully supported.
TXT	Supported	Delimited, or <code>label: value</code> format.
PDF	Beta	Goes through Document Ingestion — extracts obligations and evidence from policy documents , not transactions. Not for record-level transaction data.
.docx	Not parsed	Export to PDF or CSV first.

Column headers are matched **case-insensitively** and **language-insensitively**, with many aliases supported (including Turkish). You do not need to match capitalisation exactly.

If you upload an unsupported or empty file, ReguShield AI shows a **clear guidance message** explaining what to do — it never crashes.

3. The Four Supported Datasets

Each dataset is a distinct archetype. Use the one that matches your data. Only the required fields are mandatory; optional fields enrich the analysis and can trigger additional checks. A framework activates **only** when its fields are present.

3.1 AML / FinTech

Triggers: AMLR, AMLA 2027, AMLD6

- Required: customer_id , transaction_amount , kyc_status
- Optional: origin_country , destination_country , source_of_funds_status , pep_flag , sanctions_screening_status , transaction_count_24h , transaction_currency

Sample rows (demo values):

```
`` csv
customer_id,transaction_amount,kyc_status,origin_country,destination_country,source_of_funds_status,pep_flag,sanctions_screening_status,transaction_count_24h,transaction_currency
CUST-1001,12500,verified,DE,FR,confirmed,false,clear,3,EUR CUST-1002,48000,pending,TR,DE,unverified,true,clear,11,EUR CUST-1003,950,verified,NL,NL,confirmed,false,hit,1,EUR ``
```

Expected output: each transaction row → an AML/CTF risk-scored decision mapped to AMLR / AMLA 2027 / AMLD6 obligations, with evidence requirements for CDD/KYC gaps and recommended actions on flagged records.

3.2 Crypto / CASP

Triggers: MiCA, FATF Travel Rule

- Required: customer_id , transaction_amount , asset_type
- Optional: wallet_type , transaction_channel , customer_type , originator , beneficiary , origin_country , destination_country

Sample rows (demo values):

```
`` csv customer_id,transaction_amount,asset_type,wallet_type,transaction_channel,customer_type,originator,beneficiary,origin_country,destination_country CASP-
2001,3200,BTC,hosted,exchange,retail,Alice Demo,Bob Demo,DE,FR CASP-2002,15000,ETH,self_hosted,p2p,institutional,Acme Demo Ltd,Beta Demo GmbH,EE,DE CASP-
2003,640,USDC,hosted,exchange,retail,Carol Demo,Dan Demo,NL,NL ``
```

Expected output: each transfer row → a risk-scored decision mapped to MiCA and Travel Rule obligations (e.g. originator/beneficiary information requirements), with evidence requirements and recommended actions where data is incomplete or thresholds are crossed.

3.3 EU AI Act

Triggers: EU AI Act

- Required: ai_system_name , high_risk_ai
- Optional: ai_use_case , human_oversight , transparency_controls , training_data_controls

Sample rows (demo values):

```
`` csv ai_system_name,high_risk_ai,ai_use_case,human_oversight,transparency_controls,training_data_controls
CreditScoringEngine,true,credit_decisioning,partial,documented,reviewed ChatAssistDemo,false,customer_support,full,documented,reviewed
FraudDetectorDemo,true,aml_fraud_detection,none,missing,partial ``
```

Expected output: each AI system row → a risk-scored decision mapped to EU AI Act obligations (e.g. high-risk classification, human oversight, transparency), with evidence requirements for missing controls and recommended actions.

3.4 DORA ICT

Triggers: DORA

- Required: record_id , ict_incident
- Optional: critical_vendor , impact_level , outage_duration , recovery_status

Sample rows (demo values):

```
"" csv record_id,ict_incident,critical_vendor,impact_level,outage_duration,recovery_status ICT-3001,true,CloudHostDemo,high,4h,recovered ICT-3002,false,PaymentGwDemo,low,0,n/a
ICT-3003,true,DataCtrDemo,critical,11h,in_progress ""
```

Expected output: each ICT record row → a risk-scored decision mapped to DORA obligations (e.g. ICT risk management, incident handling), with evidence requirements for recovery/vendor gaps and recommended actions.

4. Best Practices

- **One dataset per file.** Pick the archetype that matches your data and keep it to a single file.
- **One row per record.** Every row must be self-contained — no reliance on a second file.
- **Include the required fields.** They are the minimum needed to score and map a record.
- **Add optional fields where you have them.** More signal = richer, more accurate analysis and more frameworks triggered.
- **Use a representative sample.** A few hundred realistic rows is enough to see the platform's value; you do not need your entire production dataset for the pilot.
- **Keep headers in the first row.** Aliases and translations are matched automatically — you do not need to rename columns to match exactly.
- **Prefer CSV or Excel** for record-level data; use PDF only for policy documents via Document Ingestion.

5. Common Mistakes (and How to Avoid Them)

MISTAKE	WHAT HAPPENS	FIX
Joining multiple files (customers in one, transactions in another)	The engine does not link them; records will be incomplete.	Flatten into a single file, one row per record.
Missing header row	Columns cannot be matched to fields.	Put column names in the first row.
Blank / empty file	Clear guidance message shown, nothing to analyse.	Upload a file with at least the required columns and some rows.
Wrong format (.docx)	Not parsed.	Export to PDF (for policy documents) or CSV (for records).
PDF of transactions	PDF intake is for policy documents (Document Ingestion), not transaction rows.	Put transaction data in CSV or Excel.
Missing required fields	Records may not score, and frameworks may not trigger.	Include every required field for your chosen archetype.

6. Where to Get Templates

Downloadable **sample CSVs** for each dataset are available in the in-app **Template & Data Readiness Center** on the upload page. Start from a template, replace the demo values with your data, and upload.

ReguShield AI — compliance decision-support, not legal advice. Questions: hello@regushield.ai