

Data Requirements for Preliminary Analysis

Objective:

To perform a high-quality analysis using Vernaio's Causal AI Engine, we require access to representative historical production data. Below is a compact overview of the **data types**, **format expectations**, and **transfer process**. Our team is happy to support your data engineers during preparation.

1. Required Data Domains

Please provide time-series data from the following domains:

- **Process Data:** All relevant sensor and control variables.
- **Quality Data:** All quality metrics — manual or automated.
→ *For manually entered values, please include the timestamp of measurement.*
- **Energy Data:** Preferably separated into relevant sources (electricity, steam, etc.)

If lab data or additional relevant data sources exist, they can be included.

2. Data Format Guidelines

To ensure efficient processing and minimal back-and-forth, please follow these guidelines:

Parameter	Requirement
Format	Preferred: .csv • Accepted: .parquet, .xlsx
Timestamps	ISO 8601 (e.g. 2024-05-29T13:49:00.123) • Consistent time zone
Granularity	Raw, unaggregated data at highest available resolution
Period	Data covering multiple production cycles (e.g. 1–3 months)
Data Volume	No pre-filtering needed — high-volume datasets are supported
Data Types	Supported: numeric, boolean, string
Metadata	Variable name and unit of measurement

3. Data Transfer

We provide a **secure upload link** that supports:

- **Upload-only access** (no read/download/delete possible)
- Direct delivery into our **AWS environment**
- Encrypted and access-controlled transmission

If required, we are happy to sign your **NDA** prior to data exchange.

4. Notes for Your Data Engineers

- **No preprocessing** is needed — we prefer **raw** machine data.
- Ensure **synchronization** across data domains via timestamps.
- If signals are sparse, irregular, or event-based, please include them as-is — our system supports flexible temporal structures.
- Structured data only — no PDFs, scanned documents, or charts.

Do you have any questions?

We're happy to assist your data or IT team in identifying, preparing, and transmitting the right datasets. Our goal is to make this step as frictionless as possible.