

Field Mapping Guide — UV-Vis Replicate Readings (v1.0)

Goal

Convert PDF text into rows in Output_Template.xlsx without changing any formula cells. Fill only input cells (Rep1–Rep3, Unit, Flag, Source_Location). Leave formula cells (Average, StdDev) untouched so Excel calculates automatically.

Where the data appears in the PDF

The PDF usually contains blocks like: “Sample: S-2207 Compound: API-17 Unit: AU” followed by “Replicate 1/2/3: ...”. Each Sample block becomes one output row.

Glossary

AU = absorbance unit. Replicate = repeated measurements for the same sample.

Output template contract (do not break this)

In Output_Template.xlsx, columns Average and StdDev contain Excel formulas (for example =AVERAGE(B2:D2) and =STDEV.S(B2:D2)). The agent must never overwrite these formula cells.

Column mapping

Raw text label / pattern	Meaning	Output field	How to extract	Example
Sample: S-####	Sample identifier	Sample_ID	Take the value after “Sample:”	S-2207
Compound: <name>	Compound/test name (optional)	Test_Name (optional)	Take value after “Compound:”	API-17
Unit: <unit>	Measurement unit	Unit	Take value after “Unit:”	AU
Replicate 1: <num>	Replicate #1	Rep1	Parse the number after the colon	0.812
Replicate 2: <num>	Replicate #2	Rep2	Parse the number after the colon	0.805

Replicate 3: <num>	Replicate #3	Rep3	Parse the number after the colon	0.826
-----------------------	--------------	------	--	-------

Cell writing rules (formula protection)

Allowed write targets: Rep1, Rep2, Rep3, Unit, Flag, Source_Location (and any other clearly marked input cells). Never write to Average or StdDev (or any cell that contains a formula).

Validation rules (before exporting)

Replicate values should be decimal numbers. If extracted text is missing/garbled (e.g., letters mixed in, missing decimal, or blank), leave that replicate blank, lower confidence, and flag the row.

Flag + Confidence rule of thumb

If all three replicates are present and numeric: Flag = OK and Confidence ≥ 90 . If any replicate is missing/unreadable: Flag = Needs Review and Confidence < 80 , and explain why.

Source_Location format

Use a short breadcrumb like: "Page 1 / Replicate list / Sample S-2207". If page number is unknown: "Unknown page / Replicate list / Sample S-2207".

One worked example

Raw block: "Sample: S-2207 Compound: API-17 Unit: AU ... Replicate 1: 0.812 ...
Replicate 2: 0.805 ... Replicate 3: 0.826"

Expected row: Sample_ID=S-2207; Rep1=0.812; Rep2=0.805; Rep3=0.826; Unit=AU;
Average=(leave formula); StdDev=(leave formula); Flag=OK; Source_Location=Page 1 /
Replicate list / S-2207.