

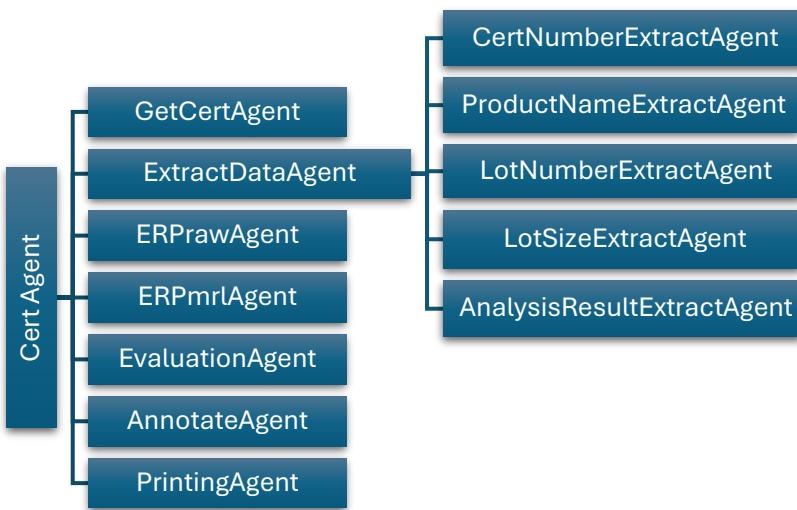
Introduction

Cert Agent V3 is an **agent-based automation system** designed to process laboratory certificate PDFs used in quality and compliance workflows (herbs & spices context).

This version (**V3**) is intentionally built as a **learning and validation system**, not a decision system. Its primary objective is to:

- Reliably extract data from **scanned paper certificates**
- Preserve extracted data **exactly as recognized**
- Enrich data using ERP reference files
- Perform **non-binding evaluations**
- Produce artifacts for **human verification**

No automatic approval, rejection, or blocking occurs in this version.



Design Philosophy

1. OCR-First by Necessity

- Certificates originate as **paper documents**
- PDFs are produced via scanning
- Embedded PDF text is unreliable or non-existent

→ the system treats every certificate as **image-based input**.

2. Single Source of Truth

- OCR is executed **once per page**
- All downstream agents consume the **same OCR output**
- No agent performs OCR independently

3. Raw Data Preservation

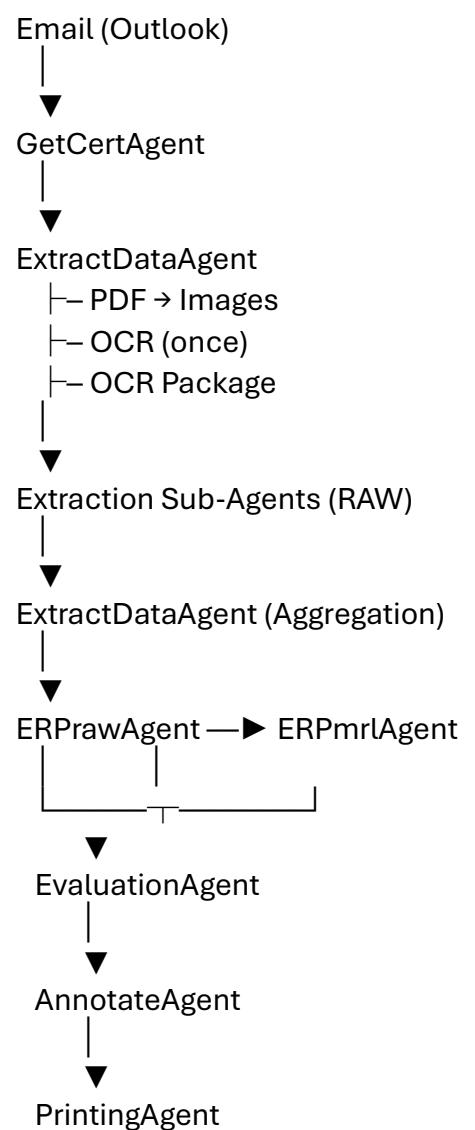
All extracted values are:

- Stored exactly as recognized
- Never corrected or normalized
- Never interpreted at extraction time

4. Separation of Concerns

Each agent:

- Has **one responsibility**
- Does **not** perform work outside its scope
- Produces explicit artifacts



❖ Folder structure

```
└── Cert-Agent-V3/
    ├── .gitattributes
    ├── .gitignore
    ├── main.py
    ├── requirements.txt
    ├── __init__.py
    └── Source_Data/
        ├── EU Limits Data.xlsx
        ├── pesticides_list.csv
        ├── products_list.csv
        └── Raw Warehouses.xlsx
    ├── PrintingAgent/
    │   ├── printer_agent.py
    │   └── __init__.py
    ├── GetCertAgent/
    │   ├── GetCertAgent ReadMe.txt
    │   └── __init__.py
    ├── ExtractDataAgent/
    │   ├── aggregator.py
    │   ├── ocr_engine.py
    │   ├── ocr_package.py
    │   ├── pdf_to_images.py
    │   ├── __init__.py
    │   ├── ProductNameExtractAgent/
    │   │   ├── product_name_agent.py
    │   │   └── __init__.py
    │   ├── LotSizeExtractAgent/
    │   │   ├── lot_size_agent.py
    │   │   └── __init__.py
    │   ├── LotNumberExtractAgent/
    │   │   ├── lot_number_agent.py
    │   │   └── __init__.py
    │   ├── CertNumberExtractAgent/
    │   │   ├── cert_number_agent.py
    │   │   └── __init__.py
    │   ├── AnalysisResultExtractAgent/
    │   │   ├── analysis_result_agent.py
    │   │   └── __init__.py
    │   ├── EvaluationAgent/
    │   │   ├── evaluator_agent.py
    │   │   └── __init__.py
    │   ├── ERPrRawAgent/
    │   │   ├── erp_raw_agent.py
    │   │   └── __init__.py
    │   ├── ERPmrIAgent/
    │   │   ├── erp_mrl_agent.py
    │   │   └── __init__.py
    ├── Docs/
    │   ├── Cert Agent info.docx
    │   ├── Cert Agent info.pdf
    │   ├── cert_agent_docs.tsx
    │   ├── Cert_Agent_V3_Full_Runbook.md
    │   ├── Cert_Agent_V3_Runbook_Step_by_Step.md
    │   ├── File Structure.txt
    │   └── ~$rt Agent info.docx
    ├── Config/
    │   ├── config.yaml
    │   ├── errors.py
    │   ├── logging.yaml
    │   ├── models.py
    │   ├── paths.py
    │   ├── utils.py
    │   └── __init__.py
    ├── AnnotateAgent/
    │   ├── annotate_agent.py
    │   └── __init__.py
    ├── .venv/
    └── .git/
```

❖ GetCertAgent

➤ Responsibility

- If Microsoft Outlook unavailable open Microsoft Outlook program, then connect to Microsoft Outlook
- If Microsoft Outlook available Connect to Microsoft Outlook
- Search Inbox for new certificate emails
- Filter emails:
- Sender: `fdesk@aldahlia.com`
- Subject contains: `CERT`, `CER` (case-insensitive)
- Download attached PDF files

➤ Input

- No external input. Operates directly on Outlook inbox.

➤ Output

- PDF files downloaded from matching emails.
- Saved to Cert Agent V3/GetCertAgent/Cert_Inbox.
- Filename: {original_pdf_name}.csv

➤ Does NOT

- Parse PDF content.
- Extract any data.
- Make any business decision.

➤ Failure behavior

- If Outlook is unavailable → wait till be available
- If no matching emails → no action → exit without error.

❖ ExtractDataAgent

➤ ExtractDataAgent is the **central spine** of Cert Agent

➤ Responsibility

1. Convert certificate PDFs into images (one image per page)
2. Perform OCR once per page.
3. Produce a shared OCR package for downstream extraction agents.
4. Coordinate extraction by invoking 5 extraction sub-agents.
5. Collect raw outputs from all sub-agents.
6. Assemble a final single CSV file with all extracted raw data per lot per certificate (for explicit multi-lot cases).
7. 5 sub agents
 - CertNumberExtractAgent
 - ProductNameExtractAgent
 - LotNumberExtractAgent
 - LotSizeExtractAgent
 - AnalysisResultExtractAgent

➤ Internal Processing

1. PDF → Images
 - One image per page
 - Preserve page order
2. OCR per page
 - Text lines
 - Line order
 - (Optional) bounding boxes
3. Build OCR Package
 - Example:
Cert Agent V3\ExtractDataAgent\Cert_To_PNG_OCR/
 - |—{original_pdf_name}_page_1.png
 - |—{original_pdf_name}_page_1_ocr.json
 - |—{original_pdf_name}_page_2.png
 - |—{original_pdf_name}_page_2_ocr.json

➤ Shared OCR Package (Contract)

1. Each {original_pdf_name}_page_X_ocr.json contains:
 - page_number
 - lines[] : text - line_index - (optional) bbox
2. That only contract is the one that all the sub-agents see.

➤ Sub-Agents Invocation

1. ExtractDataAgent:
 - Invokes each extraction sub-agent
 - Provides OCR Package only
 - Collects their CSV outputs
2. All OCR is performed only by ExtractDataAgent. Sub-agents consume OCR Package only.

➤ Aggregation Logic

1. Collect scalar outputs:
 - cert_number_raw
 - product_name_raw
 - lot_number_raw
 - lot_size_raw
2. Collect table output:
 - analyte_raw and result_raw
3. Multi-lot handling:
 - Explicit (A/B, A-B) → one CSV per lot
 - Implicit (A/2) → single CSV

➤ Input

1. PDF files stored in Cert Agent V3/GetCertAgent/Cert_Inbox

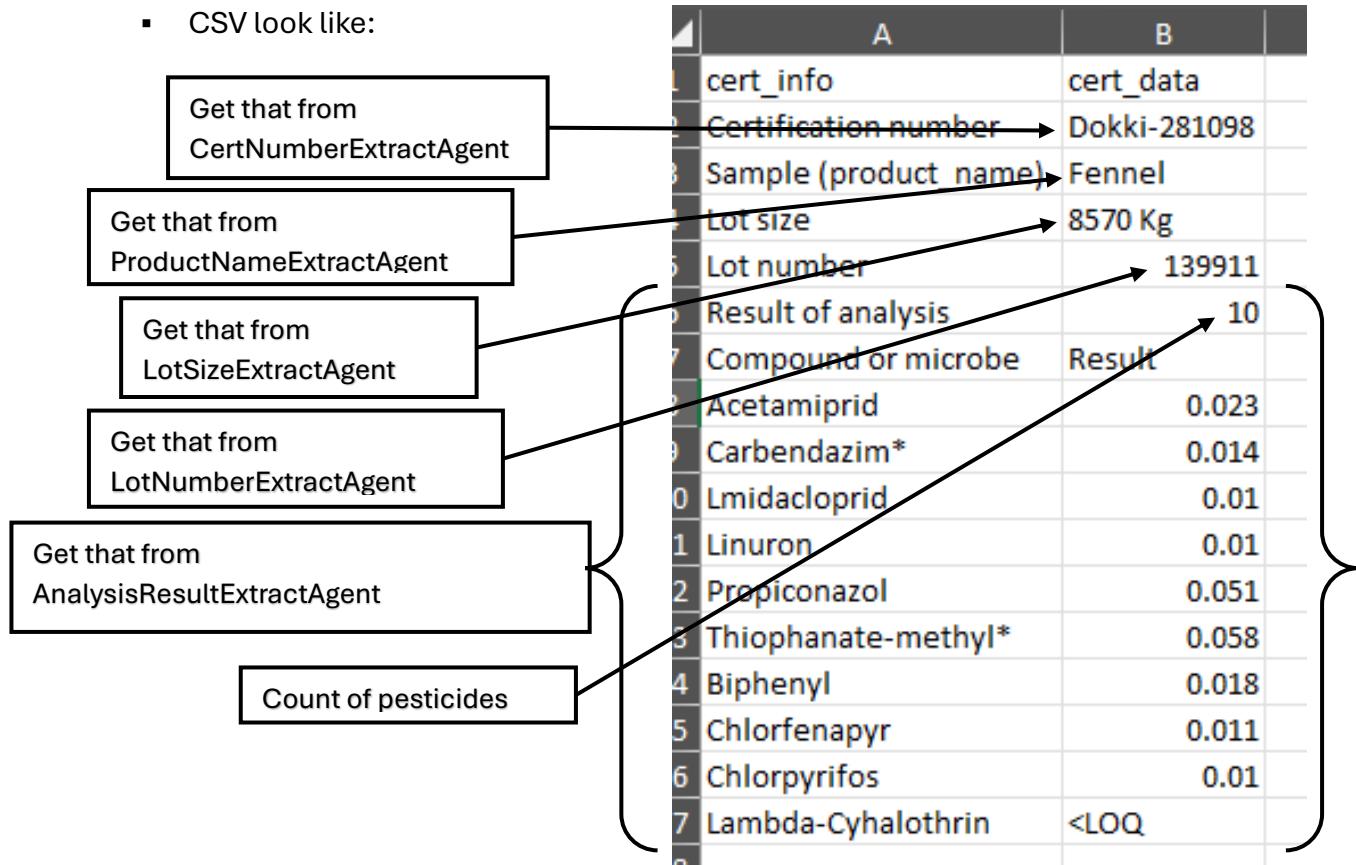
2. CSV files which stored in

- Cert Agent V3/ CertNumberExtractAgent/Cert_Number_CSVs
- Cert Agent V3/ ProductNameExtractAgent/Product_Name_CSVs
- Cert Agent V3/ LotNumberExtractAgent/Lot_Number_CSVs
- Cert Agent V3/ LotSizeExtractAgent/Lot_Size_CSVs
- Cert Agent V3/AnalysisResultExtractAgent/Analysis_Result_CSVs

➤ **Output**

1. CSV file

- One CSV per lot per certificate (for explicit multi-lot cases).
- Save it in Cert Agent V3/ ExtractDataAgent/ Cert_Data_CSVs
- Filename: {original_pdf_name}_{lot_number}_Data.csv
- If file exists → overwrite without warning
- CSV look like:



2. OCR Package

- Save it in Cert Agent V3\ExtractDataAgent\Cert_To_PNG_OCR/
- PNG Filename: {original_pdf_name}_page_X.png
- JSON Filename: {original_pdf_name}_page_X.json
- If file exists → overwrite without warning

➤ **Does NOT**

1. Parse any data meaning
2. Access ERP
3. Perform extraction logic itself.
4. Interpret values.
5. Decide business meaning.
6. Access ERP or MRL files.

➤ **Failure behavior**

1. If OCR fails on a page → page marked unreadable, pipeline continues.
2. If sub-agent output missing → corresponding fields left empty.
3. No blocking, no decisions.

❖ CertNumberExtractAgent

➤ Responsibility

- Extract raw Certificate number text from OCR package

➤ Input

- OCR Package from ExtractDataAgent stored in Cert Agent V3\ExtractDataAgent\Cert_To_PNG_OCR/

➤ Logic

- Search across all pages
- Anchor-based:
 - “Certificate Number”
 - “Cert No”
- Take nearest valid value

➤ Output

- CSV file, CSV with one row and two column: cert_info, cert_number_raw
- Save it in Cert Agent V3/CertNumberExtractAgent/Cert_Number_CSVs
- Filename: {original_pdf_name}_CertNumber.csv
- CSV look like:

	A	B
1	Cert_info	cert_number_raw
2	Certification number	Dokki-281098

➤ Does NOT

- Parse certificate number meaning
- Access ERP
- OCR
- Validate format
- Guess missing values

➤ Failure behavior

- If certificate number not found → cert_number_raw is empty.

❖ **ProductNameExtractAgent**

➤ **Responsibility**

- Extract raw Product name text from OCR Package

➤ **Input**

- OCR Package from ExtractDataAgent stored in Cert Agent V3\ExtractDataAgent\Cert_To_PNG_OCR/
- Cert Agent V3/Source_Data/ products_list.csv (that file include all product names list “if needed to get right name when extract”)

➤ **Logic**

- Search across all pages
- Anchor-based:
 - “Product name”
 - “Sample”
- Prefer header section

➤ **Output**

- CSV file, CSV with one row and two column: cert_info, product_name_raw
- Save it in Cert Agent V3/ProductNameExtractAgent/Product_Name_CSVs
- Filename: {original_pdf_name}_ProductName.csv
- CSV look like:

	A	B
1	Cert_info	product_name_raw
2	Sample (product_name)	Fennel
3		

➤ **Does NOT**

- Parse product name meaning
- Access ERP

➤ **Failure behavior**

- If product name not found → product_name_raw is empty.

❖ ***LotNumberExtractAgent***

➤ **Responsibility**

- Extract raw Lot number text from OCR Package

➤ **Input**

- OCR Package from ExtractDataAgent stored in Cert Agent
V3\ExtractDataAgent\Cert_To_PNG_OCR/

➤ **Logic**

- Search across all pages
- Anchor-based:
 - “Lot number”
- Return full raw string:
 - 139912/139913
 - 139865/2

➤ **Output**

- CSV file, CSV with one row and two columns: cert_info, lot_number_raw
- Save it in Cert Agent V3/LotNumberExtractAgent/Lot_Number_CSVs
- Filename: {original_pdf_name}_LotNumber.csv
- CSV look like:

	A	B	
1	Cert_info	lot_number_raw	
2	Lot number		139911
3			

➤ **Does NOT**

- Parse lot number meaning
- Access ERP

➤ **Failure behavior**

- If product name not found → lot_number_raw is empty.

❖ LotSizeExtractAgent

➤ Responsibility

- Extract raw Lot Size text from OCR package

➤ Input

- OCR Package from ExtractDataAgent stored in Cert Agent V3\ExtractDataAgent\Cert_To_PNG_OCR/

➤ Logic

- Search across all pages
- Anchor-based:
 - “Lot size”
- Capture value + unit as raw text

➤ Output

- CSV file, CSV with one row and one column: cert_info, lot_size_raw
- Save it in Cert Agent V3/ LotSizeExtractAgent/Lot_Size_CSVs
- Filename: {original_pdf_name}_LotSize.csv
- CSV look like:

	A	B
1	Cert_info	lot_size_raw
2	Lot size	8570 Kg
3		

➤ Does NOT

- Parse lot size meaning
- Access ERP

➤ Failure behavior

- If product name not found → lot_size_raw is empty.

❖AnalysisResultExtractAgent

➤ Responsibility

- Extract raw analysis result rows from OCR package, preserving text exactly as recognized.

➤ Input

- OCR Package from ExtractDataAgent stored in Cert Agent V3\ExtractDataAgent\Cert_To_PNG_OCR/
- Pages ordered and indexed
- Cert Agent V3/Source_Data/ pesticides_list.csv (that file include all pesticides names list “if needed to get right name when extract”)

➤ Logic

- Search across all pages
- Detect table start:
 - “Results of analysis”
 - “Compound or microbe”
- Continue across pages until section ends
- Each row:
 - analyte_raw
 - result_raw

➤ Output

- CSV file, CSV file with multiple rows and columns: analyte_raw, result_raw
- Save in Cert Agent V3/AnalysisResultExtractAgent/Analysis_Result_CSVs
- Filename: {original_pdf_name}_AnalysisResult.csv
- CSV look like:

	A	B
1	analyte_raw	result_raw
2	Result of analysis	10
3	Compound or microbe	Result
4	Acetamiprid	0.023
5	Carbendazim*	0.014
6	Lmidacloprid	0.01
7	Linuron	0.01
8	Propiconazol	0.051
9	Thiophanate-methyl*	0.058
10	Biphenyl	0.018
11	Chlorfenapyr	0.011
12	Chlorpyrifos	0.01
13	Lambda-Cyhalothrin	<LOQ
14		

➤ Does NOT

- Parse analysis result meaning
- Access ERP
- Convert results to numbers
- Compare with MRL
- Merge duplicate analytes
- Decide pass/fail
- Correct OCR mistakes
- Drop suspicious rows

هو Recorder مش Analyzer.

➤ Failure behavior

- If product name not found → result_analysis_raw is empty.
- Table anchor not found → Output CSV empty
- OCR page unreadable → Skip page, log page_number
- Partial table extracted → Output partial data (no blocking)

❖ ERPrayAgent

➤ Responsibility

- Use data extracted from certificates (lot number) to search in ERPray file (Raw Warehouses.xlsx) to get internal lot number and supplier name.
- IN ERPray file
 - Lot number in first column
 - Internal lot number in third column
 - Supplier name in forth column
 - Supplier name always written in Arabic

➤ Input

- CSVs files stored in Cert Agent V3/LotNumberExtractAgent/Lot_Number_CSVs
- ERPray file : Cert Agent V3/Source_Data/ Raw Warehouses.xlsx
- ERPray file contain multi sheets must search in all of them

➤ Output

- CSV file
- Same CSV with add new four column: cert_info, lot_number_raw ,internal_lot_number , supplier_name
- Save it in Cert Agent V3/ ERPrayAgent/ ERPray_CSVs
- Filename: {original_pdf_name}_{lot_number}_ERPray.csv
- CSV look like:

	A	B	C	D
1	cert_info	lot_number_raw	internal_lot_number	supplier_name
2	lot number	139911	lot 2601	ماهر سعد
3	الموارد التي في الشهادة موجودة في ملف ERPray في العمود رقم 1	رقم اللوشن الداخلي موجود في ملف ERPray في العمود رقم 3	اسم المورد موجود في ملف ERPray في العمود رقم 4 و لا يكتب بالعربي	

➤ Does NOT

- Decide supplier correctness
- Merge multiple suppliers
- Apply business rules

➤ Failure behavior

- If lot number not found in ERPray file → leave internal_lot_number and supplier_name empty

❖ ERPmrlAgent

➤ Responsibility

- Use data extracted from certificates (product name and result of analysis) to search in ERPmrl file (EU Limits Data.xlsx) to get limits of pesticides which found in certificate.

➤ Input

- CSVs files stored in Cert Agent V3/ ExtractDataAgent/ Cert_Data_CSVs
- ERPmrl file : Cert Agent V3/Source_Data/ EU Limits Data.xlsx
- ERPraw file contain multi sheets must search in all of them

➤ Output

- CSV file
- Same CSV with add new one column: eu_mrl
- Save it in Cert Agent V3/ ERPmrlAgent/ ERPmrl_CSVs
- Filename: {original_pdf_name}_{lot_number}_ERPMRL.csv
- CSV look like:

A	B	C
1 cert_info	cert_data	Eu_mrl
2 Certification number	Dokki-281098	
3 Sample (product_name)	Fennel	
4 Lot size	8570 Kg	
5 Lot number	139911	
6 Result of analysis	10	
7 Compound or microbe	Result	Eu_mrl
8 Acetamiprid	0.023	0.05
9 Carbendazim*	0.014	0.05
10 Lmidacloprid	0.01	0.05
11 Linuron	0.01	0.05
12 Propiconazol	0.051	0.05
13 Thiophanate-methyl*	0.058	0.05
14 Biphenyl	0.018	0.05
15 Chlufenapyr	0.011	0.05
16 Chlorpyrifos	0.01	0.05
17 Lambda-Cyhalothrin	<LOQ	0.05
18		

Get those limits
From ERPmrl file
According to product

➤ Does NOT

- Parse eu limits meaning

➤ Failure behavior

- If product not found in ERPmrl file → print product not found in ERPmrl file
- If pesticide not found in ERPmrl file → print that pesticide not found in ERPmrl file

❖ EvaluationAgent

➤ Responsibility

- Evaluate extracted laboratory results against ERP reference data.
- Perform numeric and availability comparisons only.
- Generate an evaluation report for manual review.
- Do NOT make any final decision.

➤ Input

- CSV files per (certificate × lot) from:
 - ERPrawAgent get supplier_name and internal_lot_number from: Cert Agent V3/ERPrawAgent/ERPraw_CSVs
 - ERPMrlAgent get eu_mrl from: Cert Agent V3/ERPMrlAgent/ERPMrl_CSVs
 - ExtractDataAgent get lot_number_raw and product_name_raw and analyte_raw and result_raw from: Cert Agent V3/ExtractDataAgent/Cert_Data_CSVs

➤ Output

- One Evaluation CSV file per (certificate × lot).
- Saved in: Cert Agent V3/EvaluationAgent/Evaluation_CSVs
- Filename: {original_pdf_name}_{lot_number}_Evaluation.csv
- Evaluation CSV Columns
 - cert_info,
 - lot_number,
 - supplier_name,
 - internal_lot_number,
 - analyte,
 - result_raw,
 - eu_mrl,
 - evaluation_flag,
 - evaluation_notes
- evaluation_flag values (informational only):
 - OK
 - EXCEEDED
 - MRL_NOT_FOUND
 - RESULT_NOT_NUMERIC
 - LOT_NOT_FOUND_IN_ERP

➤ Decision Rules

- ERP lookup

If internal_lot_number is empty → evaluation_flag = LOT_NOT_FOUND_IN_ERP

- Result format

If result_raw is not numeric (e.g. LOQ, ND) → evaluation_flag = RESULT_NOT_NUMERIC

- MRL availability

If eu_mrl is empty for any analyte → evaluation_flag = MRL_NOT_FOUND

- Numeric comparison

If result_raw > eu_mrl → evaluation_flag = EXCEEDED

Else → evaluation_flag = OK

➤ Does NOT

- Decide ACCEPT / REJECT / REVIEW.
- Block printing or annotation.
- Modify certificate PDFs.
- Interpret business or regulatory meaning.
- Auto-correct or guess values.

➤ Failure behavior

- If any required input CSV is missing → Generate evaluation CSV with evaluation_flag = INCOMPLETE_DATA
- If unexpected error occurs → Generate evaluation CSV with error note, pipeline continues.

AnnotateAgent

➤ Responsibility

- Annotate certificate PDFs with supplier name and internal lot number(s)

➤ Input

- PDF files stored in Cert Agent V3/GetCertAgent/Cert_Inbox
- CSV files stored in Cert Agent V3/ ERPrawAgent/ ERPraw_CSVs

➤ Output

- PDF file.
- Save it in Cert Agent V3/ AnnotateAgent/Annotated_PDFs
- Filename: {original_pdf_name}_{lot_number}_Annotated.csv
- Annotated PDF look like:

Ministry of Agriculture
Agricultural Research Center
Central Laboratory of Residue Analysis
of Pesticides and Heavy Metals in Food
QCAP

Scan for Authenticity

Test Certificate

Lot 2601

Test Certificate Details:

Certificate Number	Dokki-273859	Number of Sub Samples	N/A
Date Received	11-01-2026	Total Weight	5 Kg
Sample ID	Dokki-2026-008685	Phone	01233221799
Sample	Caraway	Fax	24175227
Variety	N/A		
Customer	Al Delta Company For Import & Export		
Address	21 Rataas St Nasr City		
Protocol Number	N/A	Number Of Packages	N/A
Lot Number	139996	Package Size	N/A Kg
Sampling Place	N/A	lot Size	1140 Kg

Analysis ID: 1402042 Start Date: 11-01-2026 End Date: 12-01-2026

Method Name: QuEChERS Method

Method Description: Quick and Easy Method (QuEChERS) for Determination of Pesticide Residues in Foods Using LC-MS/MS, GC-MS/MS European Standard Method EN 15662:2018

Results of analysis:

Compound or microbe:	Result
Cypermethrin	<LOQ
Metalaxyl	<LOQ
Pendimethalin	<LOQ
Tebuconazole	0.048 mg/kg

The measurement uncertainty expressed as expanded uncertainty (at 95% confidence level) is within the range ± 50%. The list of LOQ's attached to this certificate are listed (attachment:PR-HS-Q1, version: 7)

Person In charge: Dr. Samia Abd El Kader Elsayed

The sample was taken by the customer, the results apply to the sample as received.

Giza, Egypt Date: 12-01-2026 Thank You For Using Our Laboratory

*Marked test method are out of A2LA accreditation scope.
**These Compounds or microbe are out of A2LA accreditation scope.
This TEST CERTIFICATE shall not be reproduced except in full, without the written approval of the Laboratory.
Test results relate only to the items tested.

Page 1 of 1

7 Nadi El-said St. - Dokki - Giza - Egypt

الشارع نادي السيد - الدقى - الجزاير - مصر العربية

Dr. Hamed Abdalla

Dr. Hamed Abdalla

❖ PrintingAgent

➤ Responsibility

- Connect to local Printer
- Printer name “HP Neverstop Laser 100x”
- Print certificate annotated pdfs

➤ Input

- PDF files stored in Cert Agent V3/ AnnotateAgent/Annotated_PDFs

➤ Output

- PDF files
- Save files in" Cert Agent V3\Printed_Cert
- Filename: {original_pdf_name}_{lot_number}_Printed.csv
- If the printer not available
 - Save files in" Cert Agent V3\PrintingAgent\Waiting_To_Print_Cert

➤ Does NOT

- Parse any data meaning
- Access ERP

➤ Failure behavior

- If printer not found be calm and Retry until printer is connected (manual intervention).

❖ CertAgent

- **Responsibility**
 - Manage all other agents
- **Input**
 - .
- **Output**
 - .
- **Does NOT**
 - Contain business logic
 - Perform data extraction or comparison
- **Failure behavior**
 - If a critical agent fails (GetCertAgent / ExtractDataAgent) → stop pipeline
 - Otherwise, continue and mark missing outputs.

❖ Raw Certificate Pdf Explain

Scan for Authenticity	Test Certificate																								
Product Name Dokki Company For Import & Export Fennel N/A Al Dahlia Company For Import & Export 21 Rabaa St Nasr City	  Certification Number Dokki-281098																								
	Certificate Number : Dokki-281098 Date Received : 13-01-2026 Sample ID : Dokki-2026-010314 Sample : Fennel Variety : N/A Customer : Al Dahlia Company For Import & Export Address : 21 Rabaa St Nasr City	Number of Sub Samples : N/A Total Weight : .5 KG Phone : 0123221799 Fax : 24175227																							
	Protocol Number : N/A Lot Number : 139911 Sampling Place : N/A	Number Of Packages : N/A Packge Size : N/A Kg lot Size : 8570 Kg																							
	Analysis ID : 432368 Lot Method Name : QuEChERS Method Method Description : Quick and Easy Method (QuEChERS) for Determination of Pesticide Residues in Foods Using LC-MSMS, GC-MSMS European Standard Method EN 15662:2018 Lot weight																								
	Results of analysis : <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 40%;">Compound or microbe :</th> <th style="width: 60%;">Result</th> </tr> </thead> <tbody> <tr><td>Acetamiprid</td><td>0.023 mg/kg</td></tr> <tr><td>Carbendazim**</td><td>0.014 mg/kg</td></tr> <tr><td>Imidacloprid</td><td>0.010 mg/kg</td></tr> <tr><td>Linuron</td><td>0.010 mg/kg</td></tr> <tr><td>Propiconazol</td><td>0.051 mg/kg</td></tr> <tr><td>Thiophanate-methyl**</td><td>0.058 mg/kg</td></tr> <tr><td>Biphenyl</td><td>0.018 mg/kg</td></tr> <tr><td>Chlorfenapyr</td><td>0.011 mg/kg</td></tr> <tr><td>Chlorpyrifos</td><td>0.01 mg/kg</td></tr> <tr><td>Lambda-Cyhalothrin</td><td><LOQ</td></tr> </tbody> </table>			Compound or microbe :	Result	Acetamiprid	0.023 mg/kg	Carbendazim**	0.014 mg/kg	Imidacloprid	0.010 mg/kg	Linuron	0.010 mg/kg	Propiconazol	0.051 mg/kg	Thiophanate-methyl**	0.058 mg/kg	Biphenyl	0.018 mg/kg	Chlorfenapyr	0.011 mg/kg	Chlorpyrifos	0.01 mg/kg	Lambda-Cyhalothrin	<LOQ
	Compound or microbe :	Result																							
	Acetamiprid	0.023 mg/kg																							
	Carbendazim**	0.014 mg/kg																							
	Imidacloprid	0.010 mg/kg																							
	Linuron	0.010 mg/kg																							
Propiconazol	0.051 mg/kg																								
Thiophanate-methyl**	0.058 mg/kg																								
Biphenyl	0.018 mg/kg																								
Chlorfenapyr	0.011 mg/kg																								
Chlorpyrifos	0.01 mg/kg																								
Lambda-Cyhalothrin	<LOQ																								
Result of analysis																									
<small>The measurement uncertainty expressed as expanded uncertainty (at 95% confidence level) is within the range ± 50%. The list of LOQ's attached to this certificate are tested. (attachment: PR-HS-Q1, version 7)</small>																									
<small>Person In charge: Dr.Sanaa Abdel Kader Elsawi</small>																									
<small>The sample was taken by the customer, the results apply to the sample as received.</small>																									

الشهادات دى اصلها ورق و تعمل
الشهادات دى ممكن تبقى على صفة

لازم تبقى بالشكل ده XXXXXXX

اسم المنتج

فى ملف CSV موجود عندي فى اسمى المنتجات كلها لو محتاجينه اسمه products_list.csv

Kg ده علطول رقم بعديه

ببقي فيها اشكال كتير
الاختلافات فى رقم اللوط ممكن تبقى يالاشكال دى

1 - Lot Number : 139385

ده معناه انه لوط واحد لمورد واحد

مثال المورد سعيد حمدى لوط 2601

2- Lot Number : 139912/139913

ده معناه انه لوطين بس لمورد واحد و منتج واحد (لوطين متخللين مع بعض)

مثال : المورد ماهر سعد لوط 2601 و 2602

او اتنين موردين بس منتج واحد

مثال : المورد سعيد حمدى لوط 2601 و المورد ماهر سعد لوط 2601

3- Lot Number : 139865/2

ده معناه انه لوطين بس لمورد واحد و منتج واحد

رقم اللوط هنا هو واحد بس اللي هو 139865 و ده هيبقى مثلا ماهر سعد لوط 2605

بس لما هنكتب على الشهادة هنكتب ماهر سعد لوط 2605 1+

4- Lot Number : 139865/3

ده معناه انه 3 لوطات بس لمورد واحد و منتج واحد

رقم اللوط هنا هو واحد بس اللي هو 139865 و ده هيقي مثلا ماهر سعد لوط 2605
بس لما هنكتب على الشهادة هنكتب ماهر سعد لوط 2605 2+

5- Lot Number : 139865/5

ده معناه انه خمس لوطات بس لمورد واحد ومنتج واحد
رقم اللوط هنا هو واحد بس اللي هو 139865 و ده هيقي مثلا ماهر سعد لوط 2605
بس لما هنكتب على الشهادة هنكتب ماهر سعد لوط 2605 4+

6- Lot Number : 139679-139680

ده معناه انه لوطين بس لمورد واحد ومنتج واحد (لوطين متخللين مع بعض)
مثال : المورد ماهر سعد لوط 2601 و 2602
او اتنين موردين بس منتج واحد
مثال : المورد سعيد حمدى لوط 2601 و المورد ماهر سعد لوط 2601

7- Lot Number : SFP228

ده معناه انه لوط واحد
8- Lot Number : 163-31-03-39-2394

ده معناه انه لوط واحد
9- Lot Number : DH956-TX/2025

ده معناه انه لوط واحد
10- Lot Number : 91191

ده معناه انه لوط واحد

الشهادة اللي فيها رقمين لو دول لوطين فى شهادة واحدة
يعنى لما نبحث فى ملف الاكسل هنبحث على رقمين
مثال

Lot number : 139912/139913 او Lot number : 139912-139913

هندور على 139912 المفروض نلاقى مثلا سعيد حمدى لوط 2601
هندور على 139913 المفروض نلاقى مثلا سعيد حمدى لوط 2602
لما نكتب على الشهادة هنكتب
سعيد حمدى لوط 2601 - لوط 2602

ممکن بيقى موردين مختلفين

هندور على 139912 المفروض نلاقى مثلا سعيد حمدى لوط 2601
هندور على 139913 المفروض نلاقى مثلا ماهر سعد لوط 2611
لما نكتب على الشهادة هنكتب
سعيد حمدى لوط 2601 - ماهر سعد لوط 2611

ممکن بيقى موردين مختلفين او مورد واحد بس فيه واحد منهم مش موجود فى الاكسل

هندور على 139912 المفروض نلاقى مثلا سعيد حمدى لوط 2601
هندور على 139913 مش موجود فى الاكسل
لما نكتب على الشهادة هنكتب
سعيد حمدى لوط 2601 - N/A 139913

5. Results of analysis

فى ملف CSV موجود عندي فى اسامي المبيدات كلها لو محتاجينه اسمه pesticides_list.csv

١. أصل الشهادات وطبيعتها

كل الشهادات أصلها ورقي وتم عمل:

Scan
PDF حفظ كـ

مفيش PDF حقيقي Text-based يعتمد عليه أي Text موجود داخل الـ PDF يا إمّا OCR سابق يا إمّا reliable fragment ومش

- ## 2. عدد الصفحات
- ممكن تكون صفحة واحدة
 - ممكن تكون صفحتين (بسبب عدد المبيدات)
 - عدد الصفحات مرتبط بعدد صفوف جدول التحاليل فقط
 - الهيدر دائمًا في الصفحة الأولى
 - جدول التحاليل ممكن يبدأ في صفحة ويكمّل في الصفحة اللي بعدها بدون عنوان جديد وده أخطر جزء.

3. كل الشهادة العام (ثبتت في كل الملفات)

- كل الشهادات ليها نفس البنية المنطقية حتى لو التنسيق اختلف:
 - أعلى الصفحة الأولى (A) Header ()

- فيه دائمًا:
 - Certificate Number
 - Product Name
 - Lot Number
 - Lot Size
- والنقطة المهمة:
 - القيم دي مش في جدول
 - مكتوبة كـ Value : Label
 - أحياناً النص قريب
 - أحياناً OCR يلزفهم

(B) جدول جسم الشهادة Results of Analysis

- ده أهم جزء.

خصائص الجدول:

- عمود اسم المبيد (Compound / Microbe)
- عمود النتيجة (Result)
- ممكن يكون فيه:
 - Units ◦
 - Method ◦

شكل النتائج:

- أرقام عشرية (0.023)
- أو <LOQ
- أو LOQ
- أحياناً OCR يلخبط > أو 0/0

(C) Footer () أسفـل آخر صفحـة

- توقيع

- ختم
- تاريخ
- اسم المعمل
- ولا حاجة فيهم مهمة للاستخراج.

4. تحليل أرقام اللوت (Lot Number) من الشهادات

كل حالات اللوت اللي موثقها:

Single lot .a

- 139385 •
 - 139798 •
 - 139800 •
- ✓ straightforward

Explicit multi-lot .b

- 139912-139913 •
- د5: لوطين
- لارم search مرتين في ERP
- و يكتبهم بالشكل اللي إنت موثقه Annotation

Implicit multi-lot .c

- 139880/2 •
- د5: رقم واحد ظاهر
- implicit count
- لا ينقسم
- +1 annotation فيه
- ✓ موجود فعلياً في الشهادات

5. أسماء المنتجات (Product Name)

- كلها واضحة
- كلمة واحدة غالباً:

- Fennel ◦
- Basil ◦
- Peppermint ◦
- Marjoram ◦

لكن:

- ممكن: OCR
- يزو space
- يلخط حرف
- ProductNameExtractAgent: ف
- tolerant لازم بيقى
- raw بس يطلع

6. جدول التحاليل - أخطر استنتاج

- مفيش شهادة واحدة:

- الجدول فيها aligned perfectly
 - بعض الصفوف:
 - اسم المبيد في سطر والنتيجة تحته
 - في الصفحات الثانية:
 - مفيش Header للجدول
 - بس صفوف مباشرة
- ده يثبت إن: 
- AnalysisResultExtractAgent لازم:
- يعرف بيبدأ
 - يعرف يكمل
 - وما يعتمدش على شكل أعمدة ثابت