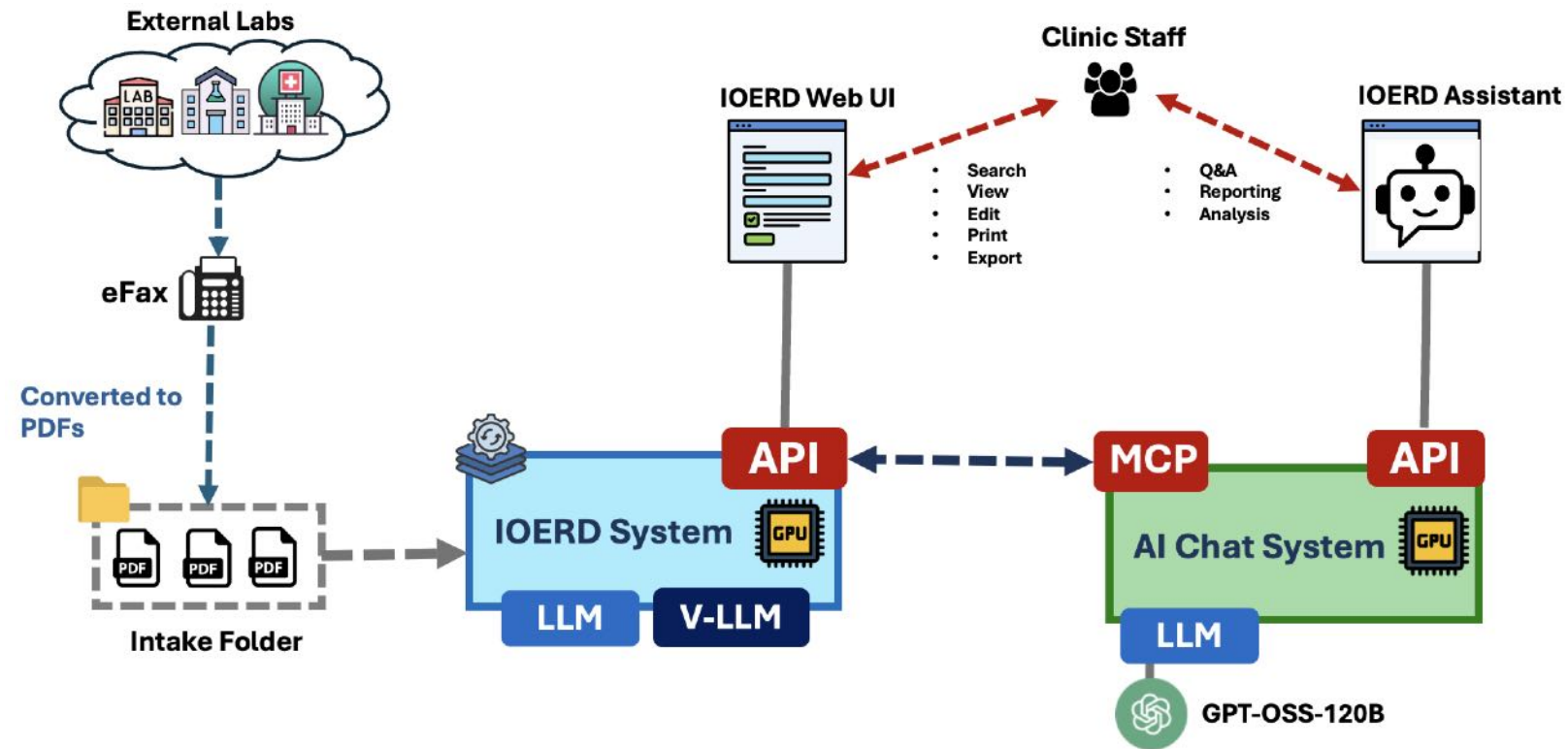


AI Powered Document Processing & Data Extraction

Including a PoC of a Medical Document Processing System (IOERD)



November 20, 2025

Robert McDermott (robert.c.mcdermott@gmail.com)

IT - Principal Architect

Agenda

AI Powered Document Processing and Data Extraction

PDF Document Basics

- Internal Document Architecture
- PDF Document Types
- Key Concepts for Data Extraction
- Data Extraction Challenges

Data Extraction Methods

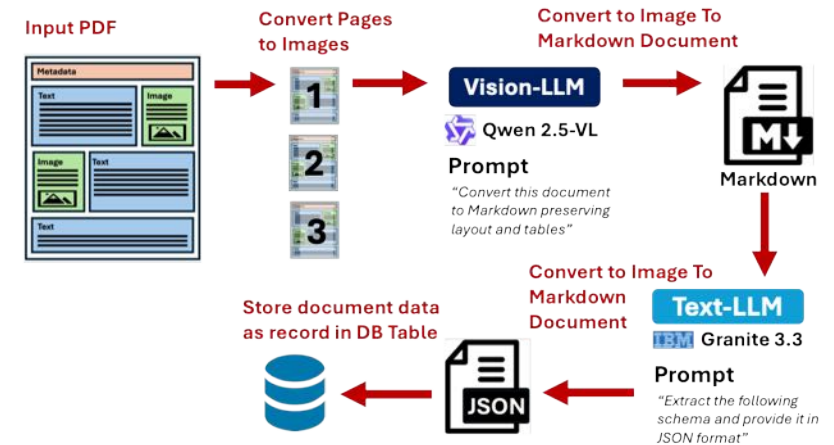
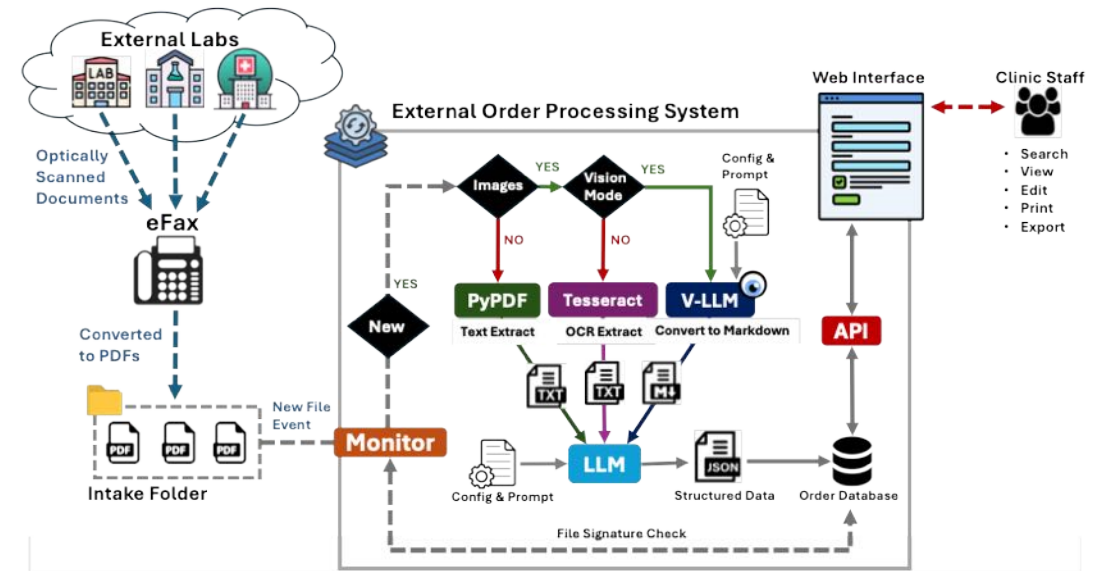
- Direct text extraction
- OCR text extraction
- Vision-LLM Text extraction and Markdown Conversion

Document to Markdown Conversion tools

- Doc2MD
- DeepSeek-OCR (Optical Context Compression)
- Docling

IOERD Clinical Fax Document Processing System PoC

- Background & Situation
- PoC Concept
- User Interface / Functionality
- API Interface for integration
- Medical document AI Assistant





PDF Document Basics

PDF Document Background

PDF Internal Architecture

File Structure (4 Main Parts)

- **Header:** Identifies the file as a PDF and its version
- **Body:** Contains the actual content objects (text, images, fonts, etc.)
- **Cross-reference table:** Index that maps object locations for quick access
- **Trailer:** Points to the cross-reference table and contains document metadata

Content Objects

- **Text objects:** Store actual character data with positioning coordinates, font information, and styling
- **Graphics objects:** Vector graphics commands (lines, curves, fills)
- **Image objects:** Raster images stored as objects, typically compressed
- **Font objects:** Embedded or referenced fonts with character mappings
- **Metadata objects:** Document properties, creation date, author, etc.

Key Concepts for Data Extraction

Text Representation

- **Native text (text-only PDFs):** Characters stored with Unicode mappings and positioning data - directly extractable
- **Image-embedded text (image-only/hybrid PDFs):** Text exists as pixels in images, not as characters - requires OCR
- PDFs use various compression schemes to reduce file size
- Each page has a content stream with drawing instructions
- Instructions are executed sequentially to render the page
- Text positioning uses coordinate systems, not semantic structure

Extraction Challenges

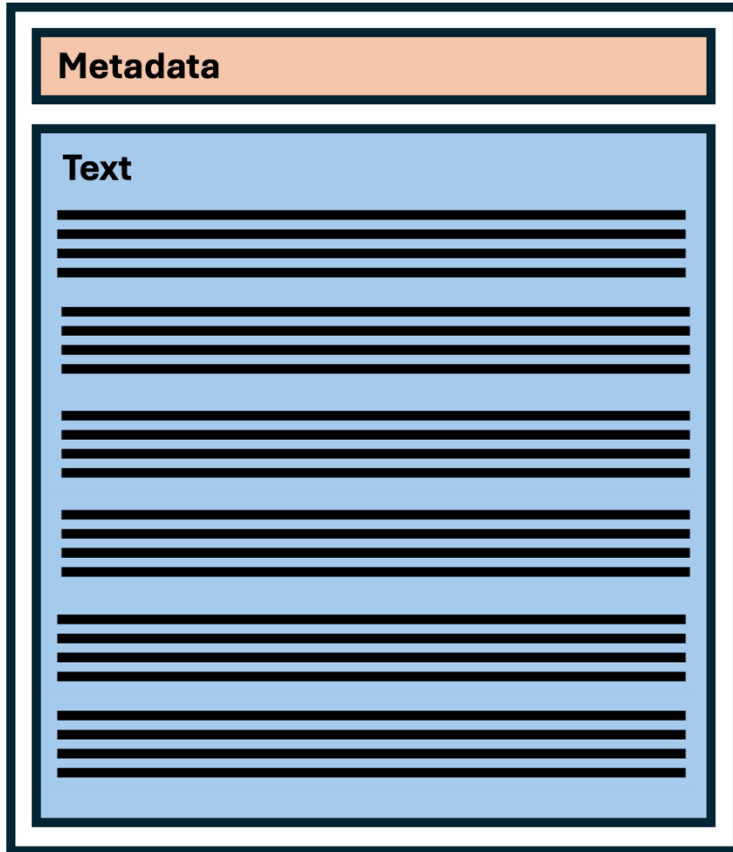
- **Text Only PDFs:** Simple text extraction using libraries
- **Hybrid PDFs:** Must extract both native text and run OCR on images to get complete content
- **Image Only PDFs (common for faxes):** Require OCR or vision-capable LLMs to extract any text

Medical Fax Implications

- Lab faxes are typically scanned documents (image only PDFs)
- May have poor quality (fax artifacts, low resolution, skewed/rotated pages)
- Traditional pipelines require image preprocessing + OCR + text extraction

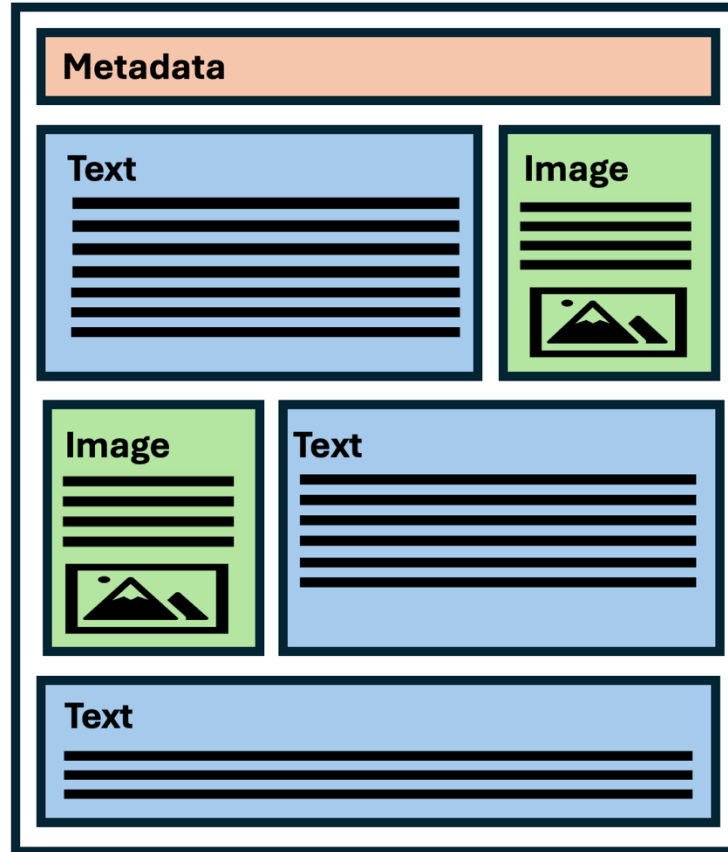
PDF Document Types

Text Only PDF



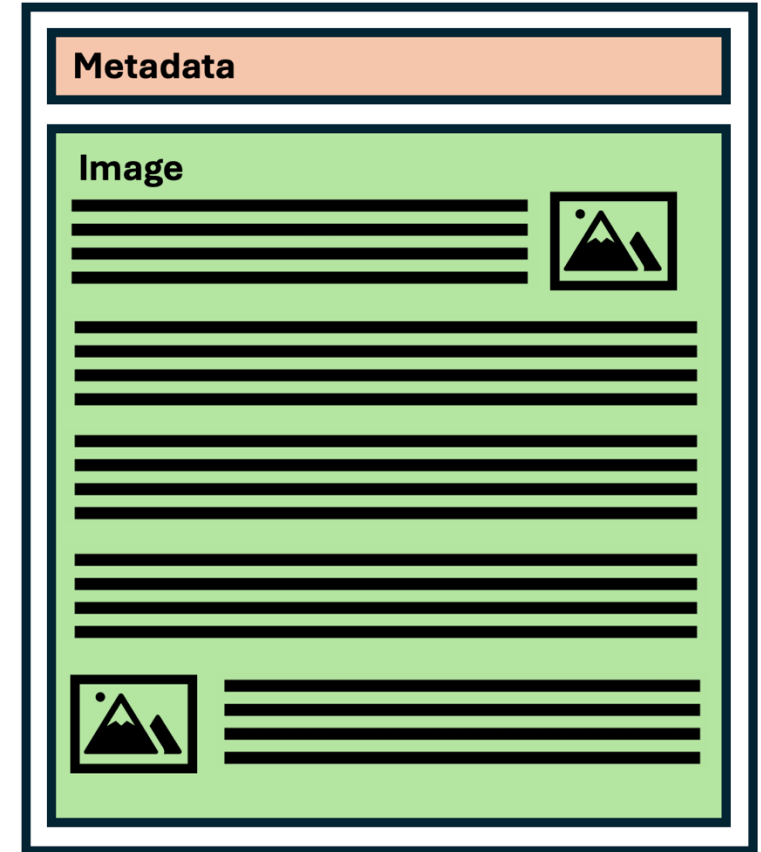
- All content directly extractable as text

Hybrid Text + Image PDF



- Mixture of text and images
- Text extraction is incomplete
- Extracting text from images requires OCR or vision capable LLM

Image Only PDF


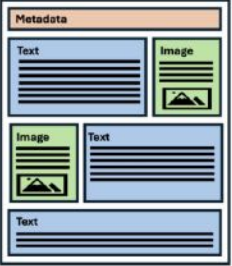



- No text is directly extractable
- Extracting text requires OCR or vision LLM
- Typical of scanned and faxed documents (medical faxes)



Extracting Information From PDF Documents


Extraction Method Results for PDF Types

PDF Type	Extraction Method	Results
Text Only 	Direct text extraction	Accurate text extraction; All text extracted; Formatting lost; Possible jumbled text
	OCR	All text extracted; Possible conversion errors; Formatting lost; Possible jumbled text;
	Vision LLM Markdown	All text extracted; Possible conversion errors; Formatting mostly preserved
Text + Images 	Direct text extraction	Text in images lost; Formatting lost; Possible jumbled text
	OCR (Tesseract)	All text extracted; Possible conversion errors; Possible jumbled text
	Vision LLM Markdown	All text extracted; Possible conversion errors; Formatting mostly preserved
Image Only 	Direct text extraction	All text lost
	OCR	All text extracted; Possible conversion errors; Formatting lost; Possible jumbled text
	Vision LLM Markdown	All text extracted; Possible conversion errors; Formatting mostly preserved



Input PDF

INVOICE



DATE August 25, 2025 **INVOICE NO** 8675309

Alien Encounters Inc
1100 Fairview Ave N.
Seattle, WA 98109
(206) 555-1234
rmcdermo@alienencounters.com

INVOICE TO
Bobby Jones
123 Main Street
Salt Lake City, UT 84118
(555) 555-0987
Bobby@gmail.com

SALESPERSON	JOB	PAYMENT TERMS	DUE DATE
Robert McDermott	Alien Sales Manager	Due on receipt	September 1, 2025

QUANTITY	DESCRIPTION	UNIT PRICE	LINE TOTAL
1	Martian Meet-and-Greet (15 minutes with a Martian ambassador)	\$150.00	\$150.00
2	Galactic Photo Session (Souvenir holo-photo with alien guest)	\$40.00	\$80.00
1	Zero-Gravity Dance with Andromedan	\$200.00	\$200.00
2	Interstellar Snack Pack (samples of alien cuisine)	\$25.00	\$50.00
3	Alien Artifact Keychain (replica meteor shard)	\$12.00	\$36.00
Subtotal			\$516.00
Sales Tax			\$51.60
Total			\$567.60



Example Code

```
# Open and read the PDF
with open(pdf_path, 'rb') as file:
    pdf_reader = pypdf.PdfReader(file)

# Extract text from each page
for page in pdf_reader.pages:
    text = page.extract_text()
    print(text)
```

- PRO:** Very Fast
- PRO:** Accurate Text
- CON:** Only works on text-based PDFs
- CON:** Formatting and structure is lost

Output Text

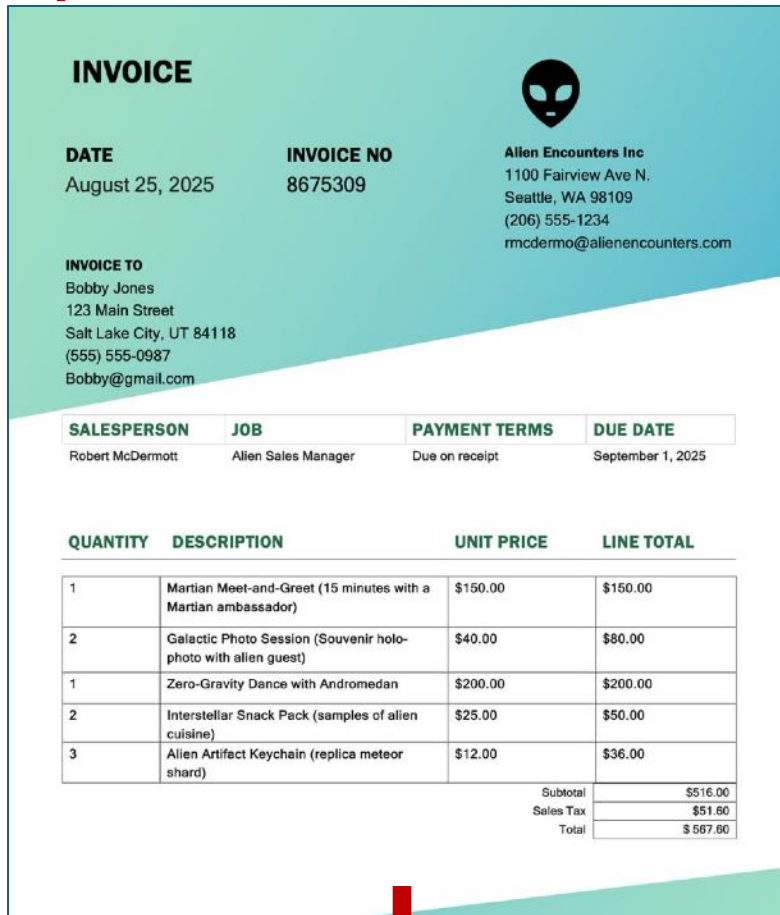
INVOICE

DATE August 25, 2025 INVOICE NO 8675309 Alien Encounters Inc 1100 Fairview Ave N. Seattle, WA 98109 (206) 555-1234 rmcdermo@alienencounters.com INVOICE TO Bobby Jones 123 Main Street Salt Lake City, UT 84118 (555) 555-0987 Bobby@gmail.com

SALESPERSON JOB PAYMENT TERMS DUE DATE Robert McDermott Alien Sales Manager Due on receipt September 1, 2025 QUANTITY DESCRIPTION UNIT PRICE LINE TOTAL 1 Martian Meet-and-Greet (15 minutes with a Martian ambassador) \$150.00 \$150.00 2 Galactic Photo Session (Souvenir holo-photo with alien guest) \$40.00 \$80.00 1 Zero-Gravity Dance with Andromedan \$200.00 \$200.00 2 Interstellar Snack Pack (samples of alien cuisine) \$25.00 \$50.00 3 Alien Artifact Keychain (replica meteor shard) \$12.00 \$36.00 Subtotal \$516.00 Sales Tax \$51.60 Total \$ 567.60



Input PDF



Tesseract

Example Code

```
# Convert PDF to images
images = convert_from_path(pdf_path, dpi=300)

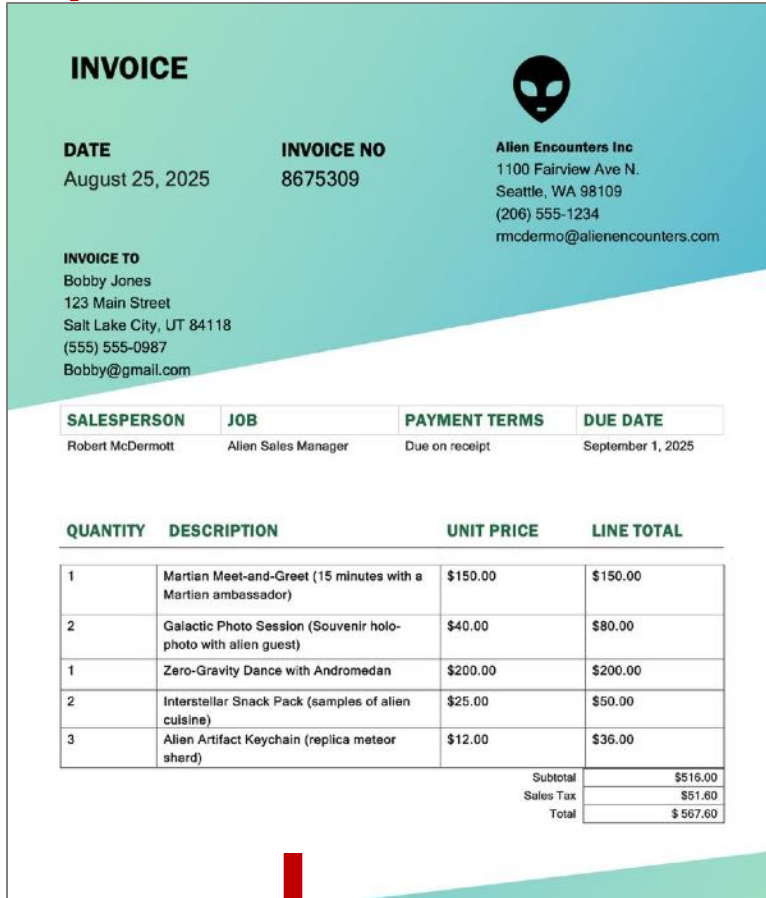
# Perform OCR on each page
for image in images:
    text = pytesseract.image_to_string(image)
    print(text)
```

- PRO:** Fast
- PRO:** Captures all text
- CON:** Formatting & structure is lost
- CON:** Text can be jumbled up
- CON:** Character recognition errors

Output Text

DATE INVOICE NO Alien Encounters Inc fo,
1100 Fairview Ave N.
August 25, 2025 8675309 Seattle, WA 98109
(206) 555-1234
rmcdermo@alienencounters.com
INVOICE TO
Bobby Jones
123 Main Street
Salt Lake City, UT 84118
(555) 555-0987
Bobby@gmail.com
SALESPERSON JOB PAYMENT TERMS | DUE DATE
Robert McDermott Alien Sales Manager Due on receipt September 1, 2025
QUANTITY DESCRIPTION UNIT PRICE LINE TOTAL
1 Martian Meet-and-Greet (15 minutes with a
Martian ambassador) | \$150.00 \$150.00
2 Galactic Photo Session (Souvenir holo- \$40.00 \$80.00
photo with alien guest)
2 Interstellar Snack Pack (samples of alien \$25.00 \$50.00
cuisine)
3 Alien Artifact Keychain (replica meteor \$12.00 \$36.00
shard)

Input PDF



INVOICE

DATE August 25, 2025 **INVOICE NO** 8675309

Alien Encounters Inc
1100 Fairview Ave N.
Seattle, WA 98109
(206) 555-1234
rmcdermo@alienencounters.com

INVOICE TO
Bobby Jones
123 Main Street
Salt Lake City, UT 84118
(555) 555-0987
Bobby@gmail.com

SALESPERSON	JOB	PAYMENT TERMS	DUE DATE
Robert McDermott	Alien Sales Manager	Due on receipt	September 1, 2025

QUANTITY	DESCRIPTION	UNIT PRICE	LINE TOTAL
1	Martian Meet-and-Greet (15 minutes with a Martian ambassador)	\$150.00	\$150.00
2	Galactic Photo Session (Souvenir holo-photo with alien guest)	\$40.00	\$80.00
1	Zero-Gravity Dance with Andromedan	\$200.00	\$200.00
2	Interstellar Snack Pack (samples of alien cuisine)	\$25.00	\$50.00
3	Alien Artifact Keychain (replica meteor shard)	\$12.00	\$36.00
Subtotal			\$516.00
Sales Tax			\$51.60
Total			\$567.60

Output Markdown

Raw Markdown

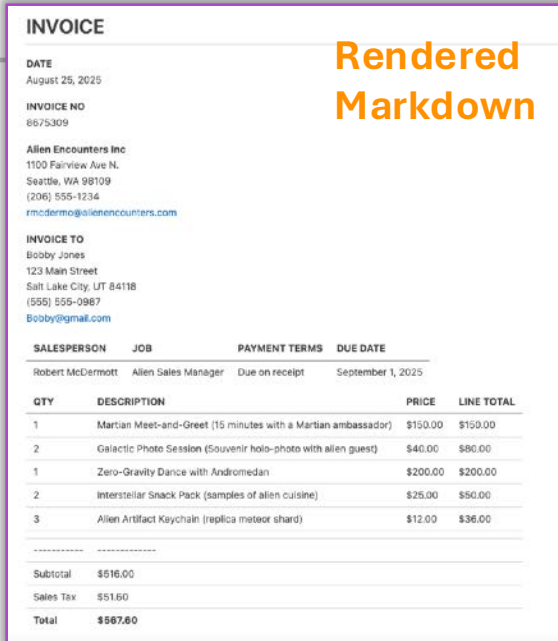
```
# INVOICE

**DATE**
August 25, 2025

**INVOICE NO**
8675309

**Alien Encounters Inc**
1100 Fairview Ave N.
Seattle, WA 98109
(206) 555-1234
rmcdermo@alienencounters.com

**INVOICE TO**
Bobby Jones
123 Main Street
Salt Lake City, UT 84118
(555) 555-0987
Bobby@gmail.com
```



INVOICE

DATE
August 25, 2025

INVOICE NO
8675309

Alien Encounters Inc
1100 Fairview Ave N.
Seattle, WA 98109
(206) 555-1234
rmcdermo@alienencounters.com

INVOICE TO
Bobby Jones
123 Main Street
Salt Lake City, UT 84118
(555) 555-0987
Bobby@gmail.com

SALESPERSON	JOB	PAYMENT TERMS	DUE DATE
Robert McDermott	Alien Sales Manager	Due on receipt	September 1, 2025

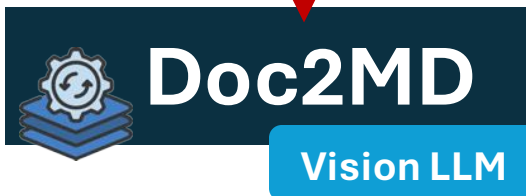
QTY	DESCRIPTION	PRICE	LINE TOTAL
1	Martian Meet-and-Greet (15 minutes with a Martian ambassador)	\$150.00	\$150.00
2	Galactic Photo Session (Souvenir holo-photo with alien guest)	\$40.00	\$80.00
1	Zero-Gravity Dance with Andromedan	\$200.00	\$200.00
2	Interstellar Snack Pack (samples of alien cuisine)	\$25.00	\$50.00
3	Alien Artifact Keychain (replica meteor shard)	\$12.00	\$36.00
Subtotal		\$516.00	
Sales Tax		\$51.60	
Total		\$567.60	

Rendered Markdown

```
| SALESPERSON | JOB | PAYMENT TERMS | DUE DATE |
|-----|-----|-----|-----|
| Robert McDermott | Alien Sales Manager | Due on receipt | September 1, 2025 |

| QTY | DESCRIPTION | PRICE | LINE TOTAL |
|-----|-----|-----|-----|
| 1 | Martian Meet-and-Greet (15 minutes with a Martian ambassador) | $150.00 | $150.00 |
| 2 | Galactic Photo Session (Souvenir holo-photo with alien guest) | $40.00 | $80.00 |
| 1 | Zero-Gravity Dance with Andromedan | $200.00 | $200.00 |
| 2 | Interstellar Snack Pack (samples of alien cuisine) | $25.00 | $50.00 |
| 3 | Alien Artifact Keychain (replica meteor shard) | $12.00 | $36.00 |

|-----|-----|-----|-----|
| Subtotal | $516.00 | | |
| Sales Tax | $51.60 | | |
| **Total** | **$567.60** | | |
```



OCR with poor quality scanned documents - Tesseract

Input PDF

Using a 10 day interval, the values in the third column will be 10 times the mean motion, n , and those in the fourth column will be 200 times the acceleration coefficient. As outlined in previous ITCG Bulletins, predictions may be made by interpolation in such a table. Its form is comparable to that of the Daily Satellite Ephemerides.

The prediction procedure involves the calculation of a time, JNL, at which the satellite is expected to appear at a particular position such as the point of local culmination. For this position, we will have calculated a value of PR. The observation then represents a measurement of the actual time at which the satellite appeared at, or very near this position. As a meticulous observer, I will have measured this time to $\pm 0^{\circ}0000001$. I will then use this actual time to recalculate the position and will correct for differences between the calculated and actual point of observation. I will also make corrections for effects of the pear shape and the ellipticity of the equator. All of this is done with the objective of obtaining a measured value of PRM that is accurate to about $\pm 0^{\circ}000001$.

Another observer, Mr. X, may either be less meticulous or may lack the means for making precise observations. He may simply use the predicted value of PRM as a measured value to correspond with his measurement of the time. We can assume that his accuracy is $\pm 0^{\circ}000001$ in timing and $\pm 0^{\circ}001$ in PRM. Let us then assume that Mr. X and I both start tracking 1960 Mu 2, using the above table and the Gear Ratio Elements. For convenience, we will also assume that his location is the same as mine and that he makes each observation at the same time that I do. After about 60 days, our records of observations might compare as shown in Table V.

TABLE V

RECORD OF OBSERVATIONS

Comparing Log of a Meticulous Observer (W.P.O.)
with Log of a Casual Observer (Mr. X)

JNL	Predicted PRM	W. P. O.		Mr. X	
		PRM	Resid.	PRM	Resid.
38462.40113315	16899.946438	16899.946518	+80	16899.9469	+400
38482.05088329	17165.472894	17165.473397	+503	17165.4732	+300
38484.04727861	17192.450126	17192.450724	+598	17192.4512	+1100
38485.08494712	17206.472112	17206.472747	+635	17206.4728	+700
38486.04369964	17219.427707	17219.428390	+683	17219.4280	+300
38494.02957430	17327.340616	17327.341718	+1102	17327.3416	+1000
38495.06723367	17341.362480	17341.363653	+1173	17341.3642	+1700
38497.06369582	17368.340068	17368.341367	+1299	17368.3410	+900

We can assume that Mr. X's values of JNL will be the same as mine (in the first column) except that they will include only five digits to the right of the decimal.

In the above table, each "residual" column represents the difference between the observed values and the predicted values, which are in the second column. As we proceed, Mr. X and I will both plot these residuals against time, as shown in Figure 2.

Tesseract

- Character Recognition Errors
- Formatting Lost

Output Text

```
| Using a 10 day interval, the values in the third column will be 10 times '
| the mean motion, n, and those in the fourth column will be 200 times the i '
| acceleration coefficient. As outlined in previous ITCG Bulletins, predictions '@' :
may be made by interpolation in such a table. Its form is comparable to that
of the Daily Satellite Ephemerides.
; | The prediction procedure involves the calculation of a time, JNL, at
which the satellite is expected to appear at a particular position such as
' the point of local culmination. For this position, we will have calculated
: a value of PR. The observation then represents a measurement of the actual
a time at which the satellite appeared at, or very near this position. As &
- meticulous observer, I will have measured this time to + 040000001, ~ will :
tL then use this actual time to recalculate the position and will correct for
7 differences between the calculated and actual point of observation. I will
i also make corrections for effects of the pear shape and the ellipticity of
So the equator. All of this is done with the objective of obtaining a measured
y value of PRY that is accurate to about + 0%00001.
. Another observer, Nr> X, may either be less meticulous or may lack the oe
' means for making precise observations. |He may simply use the predicted value
. | of PRM as a measured value to correspond with his measurement of the time. ;
We can assume that his accuracy io + 0f00001 in timing ond + 0f001 in PRU. -
Let us then assume that Mr. X and I both start tracking 1960 tu 2, using
: : the above table and the Gear Ratio Elenentse For convenience, we will also :
| assume that his location is the same as mine and that he makea each observation a,
. at the same time that I do. After about 60 days, our records of observations : 7
; ' | night compare as shoim in Table V.
an TABLE V | -
7 aS RECORD OF OBSERVATIONS
! Comparing Log of a Meticulotts Observer (W.P.O) mo,
- | with Log of a Casual Observer (Mr. X)
|
| Predicted 9 __We Pe Oe Ke X
dW) PR PR Resid. __ PR Resid. .
36462.40113315 16899.946438 16899.946518 480 16899.9469 4400
38482.05088329 17165472894 17165473397 4503 «171654752 +00 7
| 38464.04727861 17192450126 17192450724 4598 17192-4512 +1100 '
\ 58485.08494712 172060472112 17206472747 4635 = 17206.4728 +700 a
38486.04369964 17219427707 172194428390 +683 1721964280 +300 oe
|
| 5849402957450 173270340616 = 17527.341718 +1102 173527.3416 = +1000
"| 32495 .06723367 173416362480 173410565653 41175 17341.5642 41700 o0
a 38497 .06365582 17368340068 17368.541567 41299 17368.3410 +900
_ We can assume that Mre X's values of dill will be the same as mine (in the first : ;
" column) except that they will include only five digits to the right of the ;
a | decimal. '
' | In the above table, each "residual" column represents the difference .
os between the observed valuesond the predicted values, which are in the second
column. As we proceed, Mrs X and I will both plot these residuals against time, -
: as shown in Figure 26 '6 .|
. 12
```

OCR with poor quality scanned documents - Tesseract

Input PDF

TABLE 13 PRINCIPAL STATISTICS OF INDUSTRIAL PRODUCTION

Industrial Code	Industry	Number of Establishments	Number of Workers	Inputs			Work Given Out	Other Costs of Production	Output	Value Added	Sales		Employees' Remuneration	Capital Expenditure
				Total	Materials	Other Inputs					Total	Direct Exports		
				311 312	FOOD MANUFACTURING	225					9,705	482,817		
3112	Preparing and preserving of meat and meat products	5	279	3,190	3,093	97	-	362	4,222	1,032	3,871	1,376	508	261
31121	Manufacture of condensed, powdered and evaporated milk	7	416	34,811	34,170	640	75	2,477	41,956	7,070	41,641	10,233	1,967	258
31131 31132 31142	Canning and preserving of fruit and fruit juices	6	498	5,906	5,811	95	-	371	7,906	2,000	7,265	4,782	871	73
31151	Curing and preserving of fish and other sea food other than canning.. .. .	12	743	6,884	6,526	358	15	985	8,897	1,998	8,614	6,066	1,397	1,535
31152	Manufacture of coconut oil	9	338	37,577	37,041	535	39	1,235	40,192	2,577	40,275	14,100	880	122
31153	Manufacture of other refined edible oils.. .. .	6	385	52,533	51,903	631	196	2,607	56,555	3,825	55,319	26,788	1,717	4,031
31164	Processing of coffee seeds	6	299	79,808	79,676	132	149	458	84,005	4,048	84,096	37,110	1,206	1,85
31172	Manufacture of biscuits	8	884	8,917	8,467	450	85	764	12,404	3,402	12,490	718	2,046	480
31173	Manufacture of cakes and confectionery	36	557	6,246	5,908	338	1	652	8,866	2,619	8,921	-	1,143	89
31174	Manufacture of biscuits, noodles and other related products	21	333	2,739	2,624	114	-	296	4,293	1,515	4,297	11	662	60
31192 31193	Manufacture of cocoa, chocolate and sugar confectionery	20	514	2,860	2,636	224	1	796	4,566	1,705	4,529	16	884	720
31211	Manufacture of soybean products	14	854	13,924	13,372	553	-	2,084	24,626	10,701	23,559	6,644	2,590	2,285
31213	Manufacture of coffee powder	15	411	3,417	3,286	129	-	647	5,221	1,804	6,116	1,881	936	517
31215	Manufacture of ice, except dry ice	4	72	1,404	1,382	22	-	157	2,027	623	2,049	153	229	21
31216	Manufacture of cooked-food preparations	8	195	1,153	1,248	906	-	753	2,981	1,828	2,984	237	778	543
31220	Manufacture of prepared animal food	5	102	1,272	1,217	55	-	69	1,696	424	1,711	479	174	67
	Food manufacturing industries, n.e.c.	12	852	96,476	95,393	1,084	100	4,409	111,436	14,860	112,620	39,787	3,490	887
313	BEVERAGE INDUSTRIES	31	1,973	123,700	120,617	3,083	2,051	13,389	155,810	30,059	152,197	33,022	8,557	7,046
31310 31330 31341 31342 31343	Distilling and blending of spirits and brewing of beer, ale and related products	20	2,354	35,016	33,092	1,924	-	14,425	70,069	35,053	70,465	12,408	11,061	2,685
	Manufacture of soft drinks and carbonated waters	5	744	14,381	13,584	797	-	7,997	36,315	21,934	35,524	9,076	3,761	1,127
		15	1,610	20,634	19,508	1,126	-	6,429	33,754	13,119	34,941	3,332	7,300	1,558

Tesseract

Output Text

TABLE 13° PRINCIPAL STATISTICS OF INDUSTRIAL. PRODUCTION qnousand Doilors
industry
Indu trial Industry of | Number Given Costs Output Addod . : Remno- | aiture
ot
Codo Esto~ of Out of Pro- motel Dire ration
blish-| Workers Vatorials Othe? duction Exports
ments Inputs
2,553| 1839406| 30 034 | 195281

2) woop MaRSOTRNG we 205 9,705 | 462,617, 4735372| 93445 a,7i2 | 324509 | 5775621| 92092 512159 , ,
312) . | .

6 08 261
gine | Proper ond Et es °F me oe 5| 279. | 34290| 31093 97 - 362 44222, 15032 32872 1537 5 ;
sui | vomastature of condonels pondered °| ne | suen(seato| so| 5| 2877 42,956) Tx0T0| 424642|
205233| 967 |

gvop nt lic oe ee oe ee
32132} caning Sol presorving of frut we ee 6 498 5,906 5,811 95 - 371 7906 2,000 71,265 4,782 871 13
31132) | ond frolt Jemorving of fish ond 8 rT 985 g,697| 24998| 81614| 61066 1|397 | 11339
Sunde | oupng Om8 Foot other than canning» = 2r) lp 43 atstt seal 535 51 4338 | 40,292| 2017 |
40219) Bh ny oan
| atzmutoature of coconut OFF. scare ofts.- 3 ; ; 6 196 | 25607 | 561955| 31825| 951329 26,188)
1stii | 4032
3135) | omufacturo of other sere cupie otis. | - \$) FF | Zoie08 ey et 196 | ah | 845005| 4048|
84096| 37288 1206 | 185
31223 | processing of coffee te ee | 8| 8 Bioio| 8,467| 450 49| 428 | aetaoa| - 3.402| 32,490) 8)
HN 4
32172 Manufacturc oe 36 557: 6,246 5,908 338 1 652 8,866 2,619 8,921 - 13,143 89
rte Marto euro ot eres and confoatlonory ++ 22 333 25739 2,624 14 - 296 4,253| 1ad25| 42297 i 662 60
hoon, noodles an!
sum wether rointod producto, . site ont 20 514. 2,860| 2,636 224 1 196 4,566 1,705 4,529 16 884 120
31192 Manufacture of cocod, Choc _ 6 6,644) 2,590 2,285
mogot ti o « ee «e 14 854 13,924| 13,372 553 2,084 24,62 10,701 23,559 '))
558 ae ero of soya boon products «« ee 15 411 3,417 3,288 129 - "647 | - 5,221 1,804 64216 1,881
936 511
31213 Memfcture of coffeo powdor oe 4 72 1,404 1,382 22 - 157 2,027 623 2,049 153 229 2
31215 | Mermfcture of {cc, except dry ico oe 8 195 1,153 248 906 - 153 2,981) 1,828 2,984 231 718
543
31216 Manufacture of cooked-food preparationn «+ 5 102 1,272 1,217 55 - 69 1,696 424 1,711 419 LI14
61
31220 Manufacture of prepared anins) food o 12 852 96,476 95,393 1,084 100 4,409 121,436 14,860 112,
620 39,187 3,490 887
) Food manueoturing Andustrlogsy NebeCe oe 31| 1,973 J 123,700 120,617 3,083 2,052 13,389 155,810
30,059| 152197 33,022 8,557 1,046
7 3310} Distilling and blending of cpirita ond
23330 broving of boor, clo and related products 5 144 14,361| 13,584 197 - 1,997 | 36,315| 224934|
354524 _9,076| 3,762 1,121
Bre { Manufacture of soft drinks and
31343 carbonatod waters 06 «2 e# 15| 1,610 20,634| 19,508| 1,126. - 6,429 33,754| 13,219| 34,942 3,
332| 7,300 1,558

- Character recognition errors
- Formatting lost
- Jumbled text

Vision LLM with poor quality scanned documents – Doc2MD

Input PDF

Using a 10 day interval, the values in the third column will be 10 times the mean motion, n , and those in the fourth column will be 200 times the acceleration coefficient. As outlined in previous ITCF Bulletins, predictions may be made by interpolation in such a table. Its form is comparable to that of the Daily Satellite Ephemerides.

The prediction procedure involves the calculation of a time, JNL, at which the satellite is expected to appear at a particular position such as the point of local culmination. For this position, we will have calculated a value of PRM. The observation then represents a measurement of the actual time at which the satellite appeared at, or very near this position. As a meticulous observer, I will have measured this time to $\pm 0^{\circ}00000001$. I will then use this actual time to recalculate the position and will correct for differences between the calculated and actual point of observation. I will also make corrections for effects of the pear shape and the ellipticity of the equator. All of this is done with the objective of obtaining a measured value of PRM that is accurate to about $\pm 0^{\circ}000001$.

Another observer, Mr. X, may either be less meticulous or may lack the means for making precise observations. He may simply use the predicted value of PRM as a measured value to correspond with his measurement of the time. We can assume that his accuracy is $\pm 0^{\circ}000001$ in timing and $\pm 0^{\circ}0001$ in PRM. Let us then assume that Mr. X and I both start tracking 1960 Ma 2, using the above table and the Gear Ratio Elements. For convenience, we will also assume that his location is the same as mine and that he makes each observation at the same time that I do. After about 60 days, our records of observations might compare as shown in Table V.

TABLE V

RECORD OF OBSERVATIONS

Comparing Log of a Meticulous Observer (W.P.O.)
with Log of a Casual Observer (Mr. X)

JNL	Predicted PRM	W. P. O.		Mr. X	
		PRM	Resid.	PRM	Resid.
38462.40113315	16899.946438	16899.946518	+80	16899.9469	+400
38482.05088329	17165.472894	17165.473397	+503	17165.4732	+300
38484.04727861	17192.450126	17192.450724	+598	17192.4512	+1100
38485.08494712	17206.472112	17206.472747	+635	17206.4728	+700
38486.04369964	17219.427707	17219.428390	+683	17219.4280	+300
38494.02957430	17327.340616	17327.341718	+1102	17327.3416	+1000
38495.06723567	17341.362480	17341.363653	+1173	17341.3642	+1700
38497.06365582	17368.540068	17368.541367	+1299	17368.5410	+900

We can assume that Mr. X's values of JNL will be the same as mine (in the first column) except that they will include only five digits to the right of the decimal.

In the above table, each "residual" column represents the difference between the observed value and the predicted values, which are in the second column. As we proceed, Mr. X and I will both plot these residuals against time, as shown in Figure 2.



Doc2MD
Vision-LLM

- Formatting preserved; tables
- Computationally intense
- Possible hallucinations

Helpful info, but not included in the source document

Output Markdown

Rendered

Using a 10 day interval, the values in the third column will be 10 times the mean motion, n , and those in the fourth column will be 200 times the acceleration coefficient. As outlined in previous ITCF Bulletins, predictions may be made by interpolation in such a table. Its form is comparable to that of the Daily Satellite Ephemerides.

The prediction procedure involves calculating a time, JNL, at which the satellite is expected to appear at a particular position such as the points of local culmination. Based on this position, we will have calculated a value of JNL. The prediction then represents the measurement of the actual time at which the satellite appeared at, or very near this position. As a meticulous observer, I will have measured this time to ± 000000001 . I will then use this actual time to recalculate the position and will correct for differences between the calculated and actual position of observation. I will also make corrections for effects of the pear shape and the ellipticity of the equator. All of this is done with the objective of obtaining a measured value of JN that is accurate to about ± 0000001 .

Another observer, Mr. X, may either be less meticulous or may lack the means for making precise observations. He may simply use the predicted value of JN as a measured value equal to correspond with his measurement of the time. We can assume that Mr. X measures ± 0000001 times and a $1/200$ in JN. Let's imagine that Mr. X and I both start tracking 1960 Ma 2 using the above table and the Gear Ratio Elements. For convenience, we will assume that his location is the same as mine and that he makes each observation at the same time that I do. After about 60 days, our records of observations might compare as shown in Table V.

Record of Observations

Comparing Log of a Meticulous Observer (W. P. O.)

JNL	Predicted JNL	W. P. O.	Resid.*	Kr. X	Resid.*
38462.40113315	16899.946458	16899.946516	+80	16899.9469	+400
38482.0506329	17165.472894	17165.473397	+505	17165.4752	+300
38484.04727561	17192.65074	17192.650724	+598	17192.6512	+1100
38482.08494712	17206.472112	17206.472747	+635	17206.4728	+700
38484.04369984	17219.427107	17219.428390	+603	17219.4280	+300
38492.03257450	17327.540616	17327.541718	+1102	17327.5416	+1000
38495.06723567	17341.362480	17341.363635	+1173	17341.3642	+1700
38497.06365582	17368.540688	17368.541367	+1299	17368.5410	+900

*Resid. = Observed - Predicted

We can assume that Mr. X's values of JNL will be the same as mine (in the first column) except that they will include only five digits to the right of the decimal.

In the above table, each "residual" column represents the difference between the observed values and the predicted values, which are in the second column. As we proceed, Mr. X and I will both plot these residuals against time, as shown in Figure 2.

Vision LLM with poor quality scanned documents – Doc2MD

Input PDF

TABLE 13 PRINCIPAL STATISTICS OF INDUSTRIAL PRODUCTION

Industry

Thousand Dollars

Industrial Code	Industry	Number of Establishments	Number of Workers	Inputs			Work Given Out	Other Costs of Production	Output	Value Added	Sales		Employees' Remuneration	Capital Expenditure
				Total	Materials	Other Inputs					Total	Direct Exports		
311)	FOOD MANUFACTURING	225	9,705	482,817	473,372	9,445	2,712	32,509	577,621	92,091	572,553	183,406	30,034	19,181
31112	Preparing and preserving of meat and meat products	5	279	3,190	3,093	97	-	362	4,222	1,032	3,871	1,376	508	261
31121	Manufacture of condensed, powdered and evaporated milk	7	416	34,811	34,170	640	75	2,477	41,956	7,070	41,641	10,233	1,967	258
31131)	Canning and preserving of fruit and fruit juices	6	498	5,906	5,811	95	-	371	7,906	2,000	7,265	4,782	871	73
31142	Curing and preserving of fish and other sea food other than canning.. .. .	12	743	6,884	6,526	358	15	985	8,997	1,998	8,614	6,066	1,397	1,535
31151	Manufacture of coconut oil	9	338	37,577	37,041	535	39	1,235	40,192	2,577	40,275	14,100	880	122
31153	Manufacture of other refined edible oils.. .. .	6	385	52,533	51,903	631	196	2,607	56,555	3,825	55,352	26,780	1,717	185
31164	Processing of coffee seeds	6	299	79,808	79,076	732	149	458	84,005	4,048	84,096	7,225	1,056	480
31171	Manufacture of biscuits	8	804	8,917	8,467	450	85	764	12,404	3,402	12,490	37,110	2,046	480
31172	Manufacture of bread	36	927	9,646	9,628	138	7	802	8,866	2,619	8,921	-	1,143	89
31173	Manufacture of cakes and confectionery	21	333	2,739	2,624	114	-	296	4,523	1,519	4,297	11	662	60
31174	Manufacture of beehoon, noodles and other related products	20	514	2,860	2,636	224	1	796	4,566	1,705	4,529	16	884	720
31192)	Manufacture of cocoa, chocolate and sugar confectionery	14	854	13,804	13,372	553	-	2,084	24,626	10,701	23,559	6,644	2,590	2,589
31211	Manufacture of soy bean products	15	411	1,417	1,285	132	-	647	5,221	1,804	6,116	1,881	936	517
31213	Manufacture of coffee powder	5	78	1,194	1,087	96	-	157	2,027	623	2,049	153	229	21
31215	Manufacture of food flavourings	8	105	1,163	248	906	-	753	2,981	1,908	2,049	-	-	920
31216	Manufacture of cooked-food preparations	5	102	1,572	1,217	55	-	157	2,027	623	2,049	-	-	936
31290	Manufacture of prepared animal feeds	12	852	90,476	25,393	1,084	-	69	1,696	424	1,711	237	778	778
313)	FOOD MANUFACTURING INDUSTRIES, n.e.c.	31	1,973	193,700	190,437	1,083	100	4,609	111,436	14,860	112,620	39,787	3,490	887
3130)	Distilling and blending of spirits and brewing of beer, ale and related products	5	744	14,381	13,584	797	-	14,425	70,069	35,053	70,465	12,408	11,061	2,685
31310)	Manufacture of soft drinks and carbonated waters	15	1,610	20,634	19,508	1,126	-	6,429	33,754	13,119	34,941	3,332	7,300	1,558

Output Markdown

Rendered

TABLE 13 PRINCIPAL STATISTICS OF INDUSTRIAL PRODUCTION

Industry

Thousand Dollars

Industrial Code	Industry	Number of Establishments	Number of Workers	Inputs			Work Given Out	Other Costs of Production	Output	Value Added	Sales		Employees' Remuneration	Capital Expenditure
				Total	Materials	Other Inputs					Total	Direct Exports		
311)	FOOD MANUFACTURING	225	9,705	482,817	473,372	9,445	2,712	32,509	577,621	92,091	572,553	183,406	30,034	19,181
31112	Preparing and preserving of meat and meat products	5	279	3,190	3,093	97	-	362	4,222	1,032	3,871	1,376	508	261
31121	Manufacture of condensed, powdered and evaporated milk	7	416	34,811	34,170	640	75	2,477	41,956	7,070	41,641	10,233	1,967	258
31131)	Canning and preserving of fruit and fruit juices	6	498	5,906	5,811	95	-	371	7,906	2,000	7,265	4,782	871	73
31142	Curing and preserving of fish and other sea food other than canning.. .. .	12	743	6,884	6,526	358	15	985	8,997	1,998	8,614	6,066	1,397	1,535
31151	Manufacture of coconut oil	9	338	37,577	37,041	535	39	1,235	40,192	2,577	40,275	14,100	880	122
31153	Manufacture of other refined edible oils.. .. .	6	385	52,533	51,903	631	196	2,607	56,555	3,825	55,352	26,780	1,717	185
31164	Processing of coffee seeds	6	299	79,808	79,076	732	149	458	84,005	4,048	84,096	7,225	1,056	480
31171	Manufacture of biscuits	8	804	8,917	8,467	450	85	764	12,404	3,402	12,490	37,110	2,046	480
31172	Manufacture of bread	36	927	9,646	9,628	138	7	802	8,866	2,619	8,921	-	1,143	89
31173	Manufacture of cakes, confectionery	21	333	2,739	2,624	114	-	296	4,523	1,519	4,297	11	662	60
31174	Manufacture of beehoon, noodles and other related products	20	514	2,860	2,636	224	1	796	4,566	1,705	4,529	16	884	720
31192)	Manufacture of cocoa, chocolate and sugar confectionery	14	854	13,804	13,372	553	-	2,084	24,626	10,701	23,559	6,644	2,590	2,589
31211	Manufacture of soy bean products	15	411	1,417	1,285	132	-	647	5,221	1,804	6,116	1,881	936	517
31213	Manufacture of coffee powder	5	78	1,194	1,087	96	-	157	2,027	623	2,049	153	229	21
31215	Manufacture of food flavourings	8	105	1,163	248	906	-	753	2,981	1,908	2,049	-	-	920
31216	Manufacture of cooked-food preparations	5	102	1,572	1,217	55	-	157	2,027	623	2,049	-	-	936
31290	Manufacture of prepared animal feeds	12	852	90,476	25,393	1,084	-	69	1,696	424	1,711	237	778	778
313)	FOOD MANUFACTURING INDUSTRIES, n.e.c.	31	1,973	193,700	190,437	1,083	100	4,609	111,436	14,860	112,620	39,787	3,490	887
3130)	Distilling and blending of spirits and brewing of beer, ale and related products	5	744	14,381	13,584	797	-	14,425	70,069	35,053	70,465	12,408	11,061	2,685
31310)	Manufacture of soft drinks and carbonated waters	15	1,610	20,634	19,508	1,126	-	6,429	33,754	13,119	34,941	3,332	7,300	1,558



- Formatting mostly preserved; tables
- Computationally intense
- Possible hallucinations



Document to Markdown Conversion Tools

Doc2MD – PDF and Image conversion to Markdown

The first version of the IOERD document processing pipeline was using just text extraction and OCR to extract data from the medical fax documents, and it worked “OK”, but there was missing data and errors. Part of the problem was errors from the OCR process with poor quality document scans, and the other problem was pulling required data attributes from the blob of unformatted (often jumbled) text. To solve this problem, I experimented with using vision LLMs, which improved the extraction, but it still wasn’t good enough. I then discovered that if I recreated the documents in Markdown, preserving as much of the formatting as possible, especially tables, that data extraction accuracy increased dramatically. I created the Doc2MD utility (and open sourced it), which is now used as part of the IOERD PoC. Without the conversion to Markdown step prior to the structured data extraction, the accuracy of the system would not have been adequate.

 **Code Repository** <https://github.com/robert-mcdermott/doc2md>

How it works

1. Validates the input file exists and has a supported format (image or PDF)
2. For images: encodes the image in base64 format with proper MIME type detection
3. For PDFs: renders each page to a PNG image (at ~144 DPI) and processes pages sequentially
4. Constructs an OpenAI-compatible API request with structured content
5. Sends a POST request to the local API endpoint using the specified model
6. Uses the vision model to analyze the image(s) and extract text in Markdown
7. Concatenates per-page outputs for PDFs and returns a single Markdown document

Command-line arguments

- `input_path` : Path to the image or PDF file (required)
- `--model` , `-m` : Model name to use for text extraction (default: `qwen2.5vl`)
- `--endpoint` , `-e` : Endpoint URL for the OpenAI-compatible API
- `--config` , `-c` : Path to a TOML config file containing endpoint/model/api_key
- `--output` , `-o` : Output file path to write extracted text (optional, prints to stdout by default)
- `--help` , `-h` : Show help message and exit

Examples

```
# Extract text from an image using default model
uv run doc2md.py screenshot.png

# Extract text from a PDF and save to a file
uv run doc2md.py -o output.md document.pdf

# Extract text using a specific model
uv run doc2md.py --model qwen2.5vl document.jpg

# Extract text using short model flag
uv run doc2md.py -m qwen2.5vl receipt.png

# Save extracted text to a file
uv run doc2md.py --output extracted_text.md document.png

# Use custom model and save to file
uv run doc2md.py -m qwen2.5vl -o output.md invoice.jpg

# Combine all options
uv run doc2md.py --model qwen2.5vl --output result.md image.png

# Get help and see all options
uv run doc2md.py -h
```

Specifying a different API endpoint (optional)

To specify a different vision model and/or endpoint:

```
uv run doc2md.py --model <model_name> --endpoint <endpoint_url> <image_or_pdf_path>
# short flags
uv run doc2md.py -m <model_name> -e <endpoint_url> <image_or_pdf_path>
```

Authentication (optional)

Provide an API key as a Bearer token via either:

- Config file: add `api_key = "YOUR_API_KEY"` under `[llm]` (or top-level)
- Environment variable: set `DOC2MD_API_KEY` or `OPENAI_API_KEY`

Supported file types

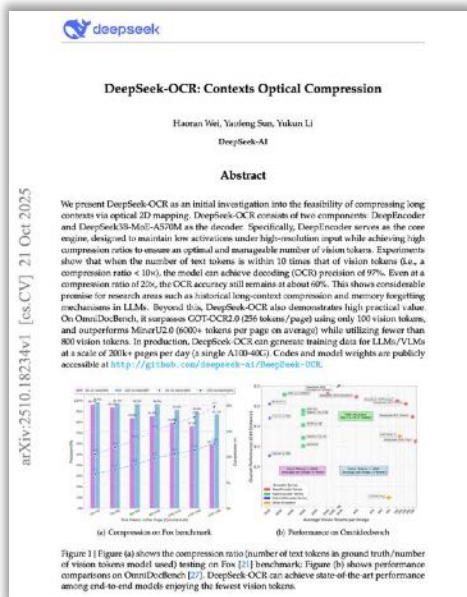
- JPG/JPEG
- PNG
- GIF
- BMP
- WebP
- PDF

DeepSeek-OCR – Optical Context Compression

DeepSeek-OCR is a breakthrough vision-language model that fundamentally reimagines document processing through what its developers call "**optical context compression.**" Rather than converting documents into thousands of text tokens, it encodes entire pages as compressed visual representations using just 100-800 vision tokens, achieving 7-20x token reduction while maintaining 97% accuracy at 10x compression. This approach leverages the insight that images can capture text, layout, and structure far more efficiently than sequential tokens, enabling a single GPU to process over 200,000 pages daily. The significance extends beyond mere efficiency: by treating documents as visual objects rather than text streams, DeepSeek-OCR addresses the fundamental long-context limitations of language models, dramatically reducing computational costs and memory requirements. This paradigm shift opens possibilities for large-scale document digitization, LLM training data generation, and real-time processing of complex documents, all while making advanced OCR capabilities accessible through its open-source architecture.

Research Paper

<https://arxiv.org/pdf/2510.18234>



Model Download

<https://huggingface.co/deepseek-ai/DeepSeek-OCR>



```
# For uv users:
uv tool install "huggingface_hub[cli]"

# For pip users:
pip install "huggingface_hub[cli]"

hf download deepseek-ai/DeepSeek-OCR --local-dir ./DeepSeek-OCR
```

Code Repository

<https://github.com/deepseek-ai/DeepSeek-OCR>



```
git clone https://github.com/DeepSeek/DeepSeek-OCR.git
```

DeepSeek-OCR – Bounding Boxes with Coordinates

Unlike the other tools we've covered, thus far, DeepSeek-OCR doesn't just output text or Markdown, it can produce multiple outputs, two of which are the original image with colored bounding boxes over each section of the document that text was extracted from, and a text/html content output that has tags and bounding box coordinates to indicate what was extracted, and from what position in the original document. This preserves information about the layout of the input document.

Output Doc with Bounding Box Overlay

INVOICE

DATE
August 25, 2025

INVOICE NO
8675309

Alien Encounters Inc
1100 Fairview Ave N.
Seattle, WA 98109
(206) 555-1234
rmcdemo@alienencounters.com

INVOICE TO
Bobby Jones
123 Main Street
Salt Lake City, UT 84118
(555) 555-0987
Bobby@gmail.com

SALESPERSON	JOB	PAYMENT TERMS	DUE DATE
Robert McDermott	Alien Sales Manager	Due on receipt	September 1, 2025

QUANTITY	DESCRIPTION	UNIT PRICE	LINE TOTAL
1	Martian Meet-and-Greet (15 minutes with a Martian ambassador)	\$150.00	\$150.00
2	Galactic Photo Session (Souvenir holo-photo with alien guest)	\$40.00	\$80.00
1	Zero-Gravity Dance with Andromedan	\$200.00	\$200.00
2	Interstellar Snack Pack (samples of alien cuisine)	\$25.00	\$50.00
3	Alien Artifact Keychain (replica meteor shard)	\$12.00	\$36.00
Subtotal			\$516.00
Sales Tax			\$51.60
Total			\$567.60

Extracted Text with Bounding Box Coordinates in the Doc

```
<|ref|>title</ref|><|det|>[[77, 52, 238, 82]]</det|>  
# INVOICE
```

```
<|ref|>text</ref|><|det|>[[68, 140, 266, 192]]</det|>  
DATE August 25, 2025
```

```
<|ref|>text</ref|><|det|>[[350, 140, 494, 192]]</det|>  
INVOICE NO 8675309
```

```
<|ref|>text</ref|><|det|>[[630, 140, 928, 247]]</det|>  
Alien Encounters Inc 1100 Fairview Ave N. Seattle, WA 98109 (206) 555- 1234 rmcdemo@alienencounters.com
```

```
<|ref|>text</ref|><|det|>[[68, 251, 295, 383]]</det|>  
INVOICE TO Bobby Jones 123 Main Street Salt Lake City, UT 84118 (555) 555- 0987 Bobby@gmail.com
```

```
<|ref|>table</ref|><|det|>[[63, 408, 933, 460]]</det|>
```

```
<table><tr><td>SALESPERSON</td><td>JOB</td><td>PAYMENT TERMS</td><td>DUE DATE</td></tr><tr><td>Robert McDermott</td><td>Alien Sales Manager</td><td>Due on receipt</td><td>September 1, 2025</td></tr></table>
```

```
<|ref|>title</ref|><|det|>[[72, 525, 352, 545]]</det|>  
# QUANTITY DESCRIPTION
```

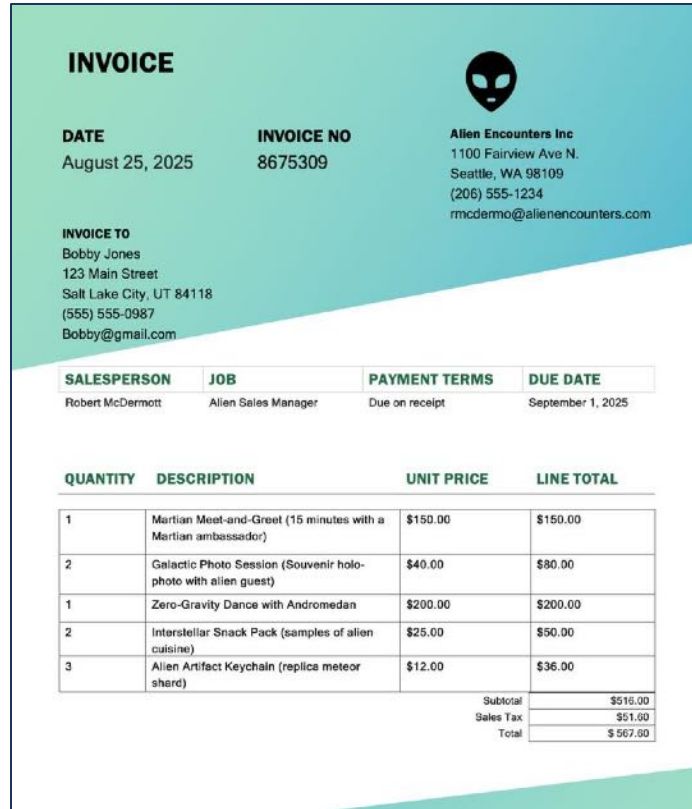
```
<|ref|>title</ref|><|det|>[[567, 525, 880, 545]]</det|>  
# UNIT PRICE LINE TOTAL
```

```
<|ref|>table</ref|><|det|>[[62, 565, 936, 824]]</det|>
```

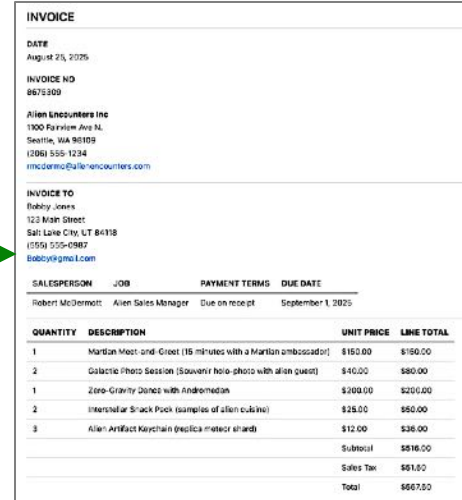
```
<table><tr><td>1</td><td>Martian Meet-and-Greet (15 minutes with a Martian ambassador)</td><td>$150.00</td><td>$150.00</td></tr><tr><td>2</td><td>Galactic Photo Session (Souvenir holo-photo with alien guest)</td><td>$40.00</td><td>$80.00</td></tr><tr><td>1</td><td>Zero-Gravity Dance with Andromedan</td><td>$200.00</td><td>$200.00</td></tr><tr><td>2</td><td>Interstellar Snack Pack (samples of alien cuisine)</td><td>$25.00</td><td>$50.00</td></tr><tr><td>3</td><td>Alien Artifact Keychain (replica meteor shard)</td><td>$12.00</td><td>$36.00</td></tr><tr><td colspan="2">Subtotal</td><td>$516.00</td><td></td></tr><tr><td colspan="2">Sales Tax</td><td>$51.60</td><td></td></tr><tr><td colspan="2">Total</td><td>$567.60</td><td></td></tr></table>
```

DeepSeek-OCR – Multiple Outputs

Input Document



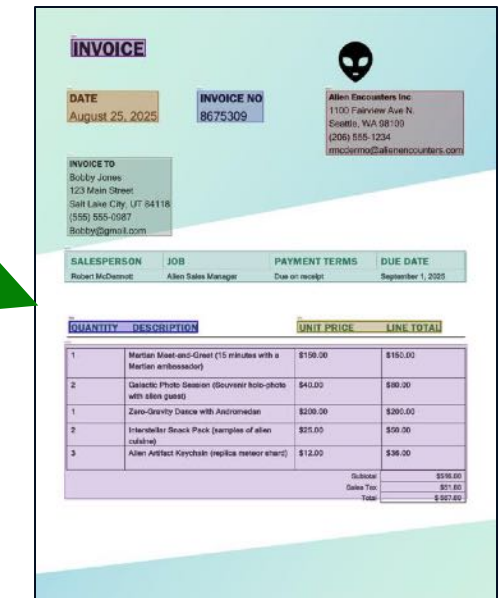
Markdown



Text with Doc Coordinates



Doc with Bounding Boxes



Example Code

```
from transformers import AutoModel, AutoTokenizer
import os, torch
model_name = 'deepseek-ai/DeepSeek-OCR'
tokenizer = AutoTokenizer.from_pretrained(model_name, trust_remote_code=True)
model = AutoModel.from_pretrained(model_name, _attn_implementation='flash_attention_2', \
    trust_remote_code=True, use_safetensors=True)
model = model.eval().cuda().to(torch.bfloat16)
prompt = "<image>\n<n|grounding>Convert the document to markdown. "
image_file = 'Alien-Encounters-Invoice.png'
output_path = 'output'
res = model.infer(tokenizer, prompt=prompt, image_file=image_file, output_path = output_path, \
    base_size = 1024, image_size = 640, crop_mode=True, save_results = True, test_compress = True)
```

Document Description

This image displays an invoice with a white background and black ink. The invoice is made out to Bobby Jones from Alien Encounters Inc. The invoice number is 8675309 and the date is August 25, 2025. The invoice lists the following items: 1 martian meet-and-greet with a martian ambassador, 1 galactic photo session with a holodeck photo, and 1 zero-gravity dance with andromeda. The total cost of the items is \$150.00. The subtotal is \$150.00. The sales tax is \$15.60. The total is \$165.60. The invoice is in a simple, easy-to-read format.

Docling



<https://github.com/docling-project/docling>

Docling simplifies document processing, parsing diverse formats, including advanced PDF understanding, and providing seamless integrations with the gen AI ecosystem.



Features

- 📁 Parsing of multiple document formats incl. PDF, DOCX, PPTX, XLSX, HTML, WAV, MP3, VTT, images (PNG, TIFF, JPEG, ...)
- 📄 Advanced PDF understanding incl. page layout, reading order, table structure, code, formulas, image classification, and more
- 🦆 Unified, expressive Docling Document representation format
- 🔄 Various export formats and options, including Markdown, HTML, DocTags and lossless JSON
- 🔒 Local execution for sensitive data and air-gapped environments
- 🤖 Plug-and-play integrations incl. LangChain, LlamaIndex, Crew AI & Haystack for agentic AI
- 🔍 Extensive OCR support for scanned PDFs and images
- 👁️ Support of several Visual Language Models (GraniteDocling)
- 🗣️ Audio support with Automatic Speech Recognition (ASR) models
- 🔑 Connect to any agent using the MCP server
- 💻 Simple and convenient CLI

Docling

Input PDF

INVOICE		INVOICE NO		Allen Encounters Inc	
DATE	August 25, 2025	INVOICE NO	8675309	1100 Fairview Ave N. Seattle, WA 98109 (206) 555-1234 rmcdermo@alienencounters.com	
INVOICE TO Bobby Jones 123 Main Street Salt Lake City, UT 84118 (555) 555-0987 Bobby@gmail.com					
SALESPERSON	JOB	PAYMENT TERMS	DUE DATE		
Robert McDermott	Alien Sales Manager	Due on receipt	September 1, 2025		
QUANTITY	DESCRIPTION	UNIT PRICE	LINE TOTAL		
1	Martian Meet-and-Greet (15 minutes with a Martian ambassador)	\$150.00	\$150.00		
2	Galactic Photo Session (Souvenir holo-photo with alien guest)	\$40.00	\$80.00		
1	Zero-Gravity Dance with Andromedan	\$200.00	\$200.00		
2	Interstellar Snack Pack (samples of alien cuisine)	\$25.00	\$50.00		
3	Alien Artifact Keychain (replica meteor shard)	\$12.00	\$36.00		
Subtotal			\$516.00		
Sales Tax			\$51.60		
Total			\$567.60		

Example Code

```
from docling.document_converter import DocumentConverter
source = "Alien-Encounters-Invoice.pdf"
converter = DocumentConverter()
result = converter.convert(source)
exports = [
    result.document.export_to_markdown(),
    result.document.export_to_html(),
    result.document.export_to_doctags()
]
for export in exports:
    print('=' * 39)
    print(export)
```

DocTags with Coordinates

```
<doctag><section_header_level_1><loc_42><loc_29><loc_124><loc_43>INVOICE</
section_header_level_1>
<text><loc_38><loc_73><loc_71><loc_81>DATE</text>
<text><loc_38><loc_87><loc_136><loc_96>August 25, 2025</text>
<text><loc_179><loc_73><loc_249><loc_81>INVOICE NO</text>
<text><loc_179><loc_87><loc_230><loc_96>8675309</text>
<section_header_level_1><loc_38><loc_128><loc_89><loc_134>INVOICE TO</
section_header_level_1>
<text><loc_38><loc_139><loc_149><loc_190>Bobby Jones 123 Main Street Salt
Lake City, UT 84118 (555) 555-0987 Bobby@gmail.com</text>
<otsl><loc_36><loc_203><loc_466><loc_229><fcel>SALESPERSON<fcel>JOB<fcel>PAYM
ENT TERMS<fcel>DUE DATE<nl><fcel>Robert McDermott<fcel>Alien Sales
Manager<fcel>Due on receipt<fcel>September 1, 2025<nl></otsl>
<section_header_level_1><loc_40><loc_264><loc_180><loc_271>QUANTITY
DESCRIPTION</section_header_level_1>
<text><loc_286><loc_264><loc_34b><loc_271>UNIT PRICE</text>
<text><loc_380><loc_264><loc_442><loc_271>LINE TOTAL</text>
<otsl><loc_35><loc_281><loc_466><loc_413><fcel>1<fcel>Martian Meet-and-Greet
(15 minutes with a Martian ambassador)<fcel>$150.00<fcel>$150.
00<nl><fcel>2<fcel>Galactic Photo Session (Souvenir holo- photo with alien
guest)<fcel>$40.00<fcel>$80.00<nl><fcel>1<fcel>Zero-Gravity Dance with
Andromedan<fcel>$200.00<fcel>$200.00<nl><fcel>2<fcel>Interstellar Snack Pack
(samples of alien cuisine)<fcel>$25.00<fcel>$50.00<nl><fcel>3<fcel>Alien
Artifact Keychain (replica meteor shard)<fcel>$12.00<fcel>$36.
00<nl><fcel>Subtotal<fcel><fcel><fcel>$516.00<nl><fcel>Sales
Tax<fcel><fcel><fcel>$51.60<nl><fcel><fcel><fcel><fcel>$ 567.60<nl></otsl>
<picture><loc_317><loc_28><loc_462><loc_124></picture>
</doctag>
```

Markdown

Rendered

INVOICE		INVOICE NO		Allen Encounters Inc	
DATE	August 25, 2025	INVOICE NO	8675309	1100 Fairview Ave N. Seattle, WA 98109 (206) 555-1234 rmcdermo@alienencounters.com	
INVOICE TO Bobby Jones 123 Main Street Salt Lake City, UT 84118 (555) 555-0987 Bobby@gmail.com					
SALESPERSON	JOB	PAYMENT TERMS	DUE DATE		
Robert McDermott	Alien Sales Manager	Due on receipt	September 1, 2025		
QUANTITY DESCRIPTION	UNIT PRICE	LINE TOTAL			
1 Martian Meet-and-Greet (15 minutes with a Martian ambassador) \$150.00 \$150.00					
2 Galactic Photo Session (Souvenir holo-photo with alien guest) \$40.00 \$80.00					
1 Zero-Gravity Dance with Andromedan \$200.00 \$200.00					
2 Interstellar Snack Pack (samples of alien cuisine) \$25.00 \$50.00					
3 Alien Artifact Keychain (replica meteor shard) \$12.00 \$36.00					
Subtotal					\$516.00
Sales Tax					\$51.60
					\$ 567.60

HTML

Rendered

INVOICE		INVOICE NO		Allen Encounters Inc	
DATE	August 25, 2025	INVOICE NO	8675309	1100 Fairview Ave N. Seattle, WA 98109 (206) 555-1234 rmcdermo@alienencounters.com	
INVOICE TO Bobby Jones 123 Main Street Salt Lake City, UT 84118 (555) 555-0987 Bobby@gmail.com					
SALESPERSON	JOB	PAYMENT TERMS	DUE DATE		
Robert McDermott	Alien Sales Manager	Due on receipt	September 1, 2025		
QUANTITY DESCRIPTION	UNIT PRICE	LINE TOTAL			
1 Martian Meet-and-Greet (15 minutes with a Martian ambassador) \$150.00 \$150.00					
2 Galactic Photo Session (Souvenir holo- photo with alien guest) \$40.00 \$80.00					
1 Zero-Gravity Dance with Andromedan \$200.00 \$200.00					
2 Interstellar Snack Pack (samples of alien cuisine) \$25.00 \$50.00					
3 Alien Artifact Keychain (replica meteor shard) \$12.00 \$36.00					
Subtotal					\$516.00
Sales Tax					\$51.60
					\$ 567.60

IOERD Fax Processing System PoC

IOERD Fax Processing System PoC

Background

IOERD stands for "Internally Ordered External Results Delivery," and is a workflow concept used in clinics and hospitals where labs, and other medical services are requested internally by clinicians that are fulfilled by external sources, such as outside laboratories, or third-party healthcare providers.

Situation

Clinics can receive several hundreds (even thousands) of faxes from external providers such as genetic testing, bloodwork results, pathology reports, radiology and other. The faxes come in via an eFax service and are delivered as PDFs. Clinic staff must manually review the faxes to determine what patient the results belong to and then attach the PDFs to the media tab in the patients record in the Epic system.

Opportunity

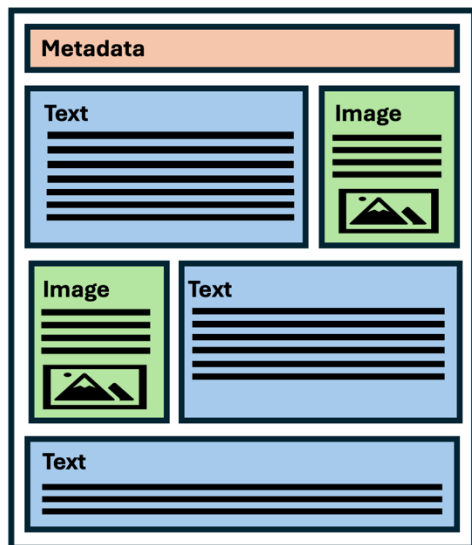
Electronically process new IOERD documents as they come in, extract the identifying information that clinic staff can monitor what orders have been delivered, with search capabilities to make identifying the orders more efficient. Phase 1 is stand-alone and not integrated with the EHR system. Future phases could potentially include the ability for staff to review the order in the processing system and make any adjustments as required and then approve patient recorded attachments in the EHR. Perhaps after running in this human-in-the-loop mode for some period to refine the system and gain confidence, incoming external results could automatically update patient records.

Status

A minimal working PoC has been created to validate the capability of such as system. The system functions as intended but has only been run in a development mode, and it not part of any clinic process or integrated with any production systems.

IOERD Fax Processing System PoC – High-Level Concept

Input PDF



Convert Pages to Images



Convert to Image To Markdown Document



Prompt

“Convert this document to Markdown preserving layout and tables”



Markdown

Convert to Image To Markdown Document

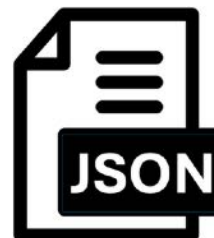
Text-LLM

IBM Granite 3.3

Prompt

“Extract the following schema and provide it in JSON format”

Store document data as record in DB Table

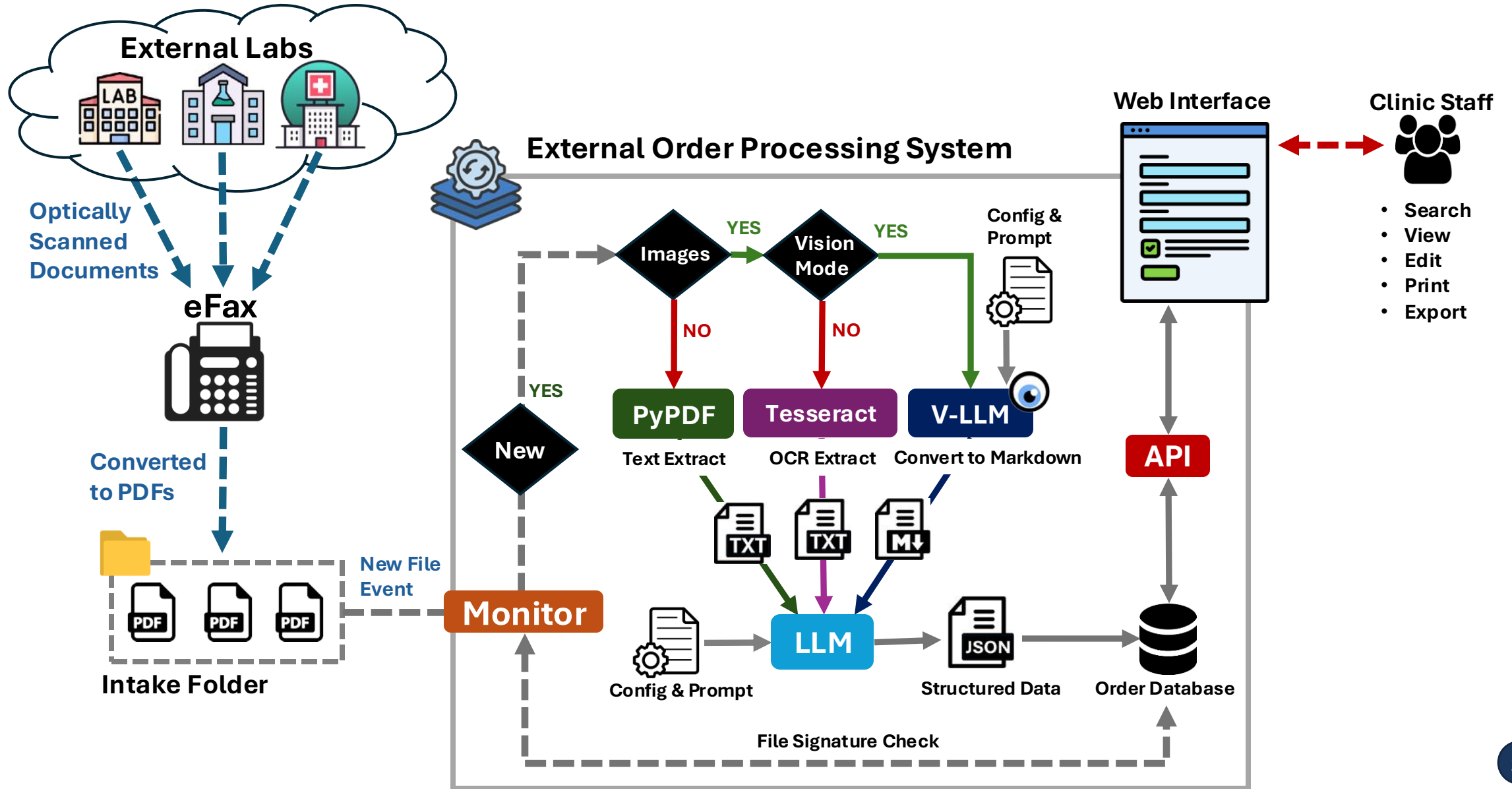


IOERD Fax Processing System PoC - Data Extraction Schema

A data schema must be defined in the system configuration file for each pipeline instance and document type. This schema is included in the prompt for the data extraction phase of the pipeline.

Name	patient_full_name	specimen_number
Type	String	String
Description	Complete legal name of the patient as it appears on identification documents	Date when the specimen was collected from the patient in MM/DD/YYYY format
Name	patient_identifier	collection_date
Type	String	Date
Description	Unique patient ID, medical record number, or other identifying number assigned by the healthcare facility	Date when the specimen was collected from the patient in MM/DD/YYYY format
Name	patient_address	test_name
Type	Text	String
Description	Complete mailing address of the patient including street, city, state, and ZIP code	Name or description of the laboratory test or panel that was performed
Name	patient_dob	result_date
Type	Date	Date
Description	Patient's date of birth in MM/DD/YYYY format	Date when the test results were finalized and reported in MM/DD/YYYY format
Name	ordering_provider_name	performing_lab
Type	String	String
Description	Full name of the physician or healthcare provider who ordered the laboratory test	Name of the laboratory facility that performed the testing and analysis

External Delivered Medical Fax Processing System PoC



IOERD Fax Processing System PoC – Medical Lab Results Documents

For this presentation, the following 20 lab result reports were copied into the intake folder to start the pipeline. *These reports contain no PHI; all information in them is fabricated.*

Pacific Diagnostic Center
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.pacificdiagnostic.com

PATIENT INFORMATION
Patient Name: Julia S. Anderson Date of Birth: 02/12/1981
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Henry J. Olson, MD Collection Date: 10/25/2023
Test Ordered: Exchange Estimation Blank (E56) Result Date: 10/26/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Turbidity	0	units	0-10		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Tri-State Endocrine Labs
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.tristateendocrine.com

PATIENT INFORMATION
Patient Name: Carl Anderson Date of Birth: 12/05/1962
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Grant S. Nelson, MD Collection Date: 11/02/2023
Test Ordered: Creatinine (Cr) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Cr (Serum)	0.8	mg/dL	0.6-1.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Cascade Pathology Group
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.cascadepathology.com

PATIENT INFORMATION
Patient Name: Robert J. Thomas Date of Birth: 03/15/1954
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Jonathan B. Bell, MD Collection Date: 11/02/2023
Test Ordered: C-Reactive Protein (CRP) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
CRP (hs)	0.2	mg/L	0-1.0		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

SoundLab Diagnostics
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.soundlab.com

PATIENT INFORMATION
Patient Name: Mark T. Campbell Date of Birth: 03/07/1961
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Arnette G. Long, MD Collection Date: 03/07/2024
Test Ordered: BNP (B-type Natriuretic Peptide) Result Date: 03/07/2024

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
BNP (hs)	88	ng/mL	<100		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

SoundLab Diagnostics
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.soundlab.com

PATIENT INFORMATION
Patient Name: Robert J. Thomas Date of Birth: 03/15/1954
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert J. Thomas, MD Collection Date: 11/02/2023
Test Ordered: Hemoglobin (Hgb) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Hgb	15.1	g/dL	13.8-16.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Northwest Clinical Laboratories
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.nwclinical.com

PATIENT INFORMATION
Patient Name: Robert J. Thomas Date of Birth: 03/15/1954
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert J. Thomas, MD Collection Date: 11/02/2023
Test Ordered: Hemoglobin (Hgb) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Hgb	15.1	g/dL	13.8-16.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Pacific Diagnostic Center
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.pacificdiagnostic.com

PATIENT INFORMATION
Patient Name: Robert J. Thomas Date of Birth: 03/15/1954
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert J. Thomas, MD Collection Date: 11/02/2023
Test Ordered: Hemoglobin (Hgb) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Hgb	15.1	g/dL	13.8-16.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Tri-State Endocrine Labs
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.tristateendocrine.com

PATIENT INFORMATION
Patient Name: Daniel S. King Date of Birth: 08/17/1982
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Anthony G. Manno, MD Collection Date: 11/02/2023
Test Ordered: Serum TSH Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
TSH	1.2	uIU/mL	0.4-4.0		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Pacific Diagnostic Center
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.pacificdiagnostic.com

PATIENT INFORMATION
Patient Name: Michael S. Baker Date of Birth: 01/21/1953
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: John E. Olson, MD Collection Date: 10/25/2023
Test Ordered: Serum Creatinine Test Result Date: 10/25/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Cr (Serum)	0.8	mg/dL	0.6-1.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

SoundLab Diagnostics
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.soundlab.com

PATIENT INFORMATION
Patient Name: Thomas S. Hill Date of Birth: 07/22/1977
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Kevin D. Johnson, MD Collection Date: 10/25/2023
Test Ordered: Serum Creatinine Test Result Date: 10/25/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Cr (Serum)	1.8	mg/dL	0.7-1.3		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Cascade Pathology Group
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.cascadepathology.com

PATIENT INFORMATION
Patient Name: Daniel S. King Date of Birth: 08/17/1982
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Kevin D. Johnson, MD Collection Date: 10/25/2023
Test Ordered: Serum Creatinine Test Result Date: 10/25/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Cr (Serum)	1.8	mg/dL	0.7-1.3		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Northwest Clinical Laboratories
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.nwclinical.com

PATIENT INFORMATION
Patient Name: Robert J. Thomas Date of Birth: 03/15/1954
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert J. Thomas, MD Collection Date: 11/02/2023
Test Ordered: Hemoglobin (Hgb) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Hgb	15.1	g/dL	13.8-16.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Cascade Pathology Group
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.cascadepathology.com

PATIENT INFORMATION
Patient Name: Robert J. Thomas Date of Birth: 03/15/1954
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert J. Thomas, MD Collection Date: 11/02/2023
Test Ordered: Hemoglobin (Hgb) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Hgb	15.1	g/dL	13.8-16.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Northwest Clinical Laboratories
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.nwclinical.com

PATIENT INFORMATION
Patient Name: Robert M. Gentry Date of Birth: 06/21/1958
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert M. Gentry, MD Collection Date: 10/25/2023
Test Ordered: Prostate Specific Antigen (PSA) Test Result Date: 10/25/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
PSA (Total)	2.4	ng/mL	0.0-4.0		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Tri-State Endocrine Labs
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.tristateendocrine.com

PATIENT INFORMATION
Patient Name: Patrick W. Kim Date of Birth: 08/13/1990
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Louis H. Brown, MD Collection Date: 10/25/2023
Test Ordered: C-Reactive Protein, High Sensitivity (hs-CRP) Result Date: 10/25/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
hs-CRP	1.4	mg/L	<3.0		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Tri-State Endocrine Labs
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.tristateendocrine.com

PATIENT INFORMATION
Patient Name: Laura S. Williams Date of Birth: 07/14/1959
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert M. Gentry, MD Collection Date: 11/02/2023
Test Ordered: Thyroid Stimulating Hormone (TSH) Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
TSH	0.4	uIU/mL	0.4-4.0		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Northwest Clinical Laboratories
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.nwclinical.com

PATIENT INFORMATION
Patient Name: Felix D. Ortega Date of Birth: 02/12/1958
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Gregory W. Adams, MD Collection Date: 11/02/2023
Test Ordered: Vitamin D, 25-Hydroxy Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
25-OH Vitamin D	18	ng/mL	30-60		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

SoundLab Diagnostics
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.soundlab.com

PATIENT INFORMATION
Patient Name: Sara J. Johnson Date of Birth: 01/12/1988
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert J. Thomas, MD Collection Date: 11/02/2023
Test Ordered: Vitamin D, 25-Hydroxy Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
25-OH Vitamin D	18	ng/mL	30-60		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Cascade Pathology Group
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.cascadepathology.com

PATIENT INFORMATION
Patient Name: Robert J. Thomas Date of Birth: 03/15/1954
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert J. Thomas, MD Collection Date: 11/02/2023
Test Ordered: Hemoglobin (Hgb) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
Hgb	15.1	g/dL	13.8-16.2		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

Pacific Diagnostic Center
12400 Pacific Blvd, Seattle, WA 98148
Phone: (206) 461-1100 Fax: (206) 461-1101
www.pacificdiagnostic.com

PATIENT INFORMATION
Patient Name: Lisa R. Williams Date of Birth: 10/24/1985
Patient ID: F94866322 Specimen Number: ACC-23-0912309

TEST INFORMATION
Ordering Provider: Robert M. Gentry, MD Collection Date: 11/02/2023
Test Ordered: Prostate Specific Antigen (PSA) Test Result Date: 11/02/2023

LABORATORY RESULTS

Test Name	Result	Unit	Reference Range	Flag	Notes
PSA (Total)	1.8	ng/mL	0.0-4.0		Normal

Electronically Signed By: Medical Director Dr. James H. Patterson, MD, PhD
Signed: 1/25/2025 14:32 PST
This report has been electronically signed and does not require a handwritten signature.

IOERD Fax Processing System PoC – User Interface

Administration and Monitoring Interface

Total Documents 20

Successful 20

Failed 0

Success Rate 100.0%

File Monitoring

Status: Active

Files Detected: 20

Files Processed: 20

Processing Errors: 0

Start Monitoring

Stop Monitoring

Manual Scan

Processing Methods

Text Extraction: 0	OCR Fallback: 0
Hybrid Method: 0	Vision Processing: 20

Vision Processing

Status: Available

Success Rate: 100.0%

Total Processed: 20

Fallback Rate: 0.0%

Performance

Total Processing Time: 1200.00s

Average Time per Document: 60.00s

Queue Size: 0

Uptime: 0s

Component Status

Text Extractor ✓ Ocr Engine ✓ Llm Processor ✓

Vision Processor ✓

System Information

Configuration	Components
Schema Fields: 10	Text Extractor: Available (pypdf)
Database Path: documents.db	Ocr Engine: Available (tesseract)
Intake Folder: intake	Llm Processor: Available - granite3.3:latest
	Vision Processor: Available
	Model: qwen2.5vl:latest
	Endpoint: http://localhost:11434/v1/chat/completions

IOERD Fax Processing System PoC - User Interface

Search Interface

Search Filters

Text Search

Patient Full Name

Patient Identifier

Patient Address

Patient Dob

Ordering Provider Name

Specimen Number

Collection Date

Test Name

Result Date

Performing Lab

Sort Order

Sort By

Sort Order

Search

Clear

Search Results

20 documents

Filename	Processed	Method	Patient Full Name	Patient Identifier	Patient Address	Patient Dob	Ordering Provider Name	Specimen Number	Collection Date	Test Name	Result Date	Performing Lab	Actions
Lab Report - Trevor S. Holt.pdf	11/16/2025	Vision	Trevor S. Holt	FH-00601467	6214 Briarwood Pl, Bellevue, WA 98006	1977-07-30	Karen D. Ishikawa, MD	ACC-23-0912356	2025-11-01	Serum Magnesium	2025-11-01	SoundLab Diagnostics	View
Lab Report - Sofia N. Morales.pdf	11/16/2025	Vision	Sofia N. Morales	FH-00660322	2142 Highland Trail, Seattle, WA 98112	1989-02-23	Henry J. Olson, MD	ACC-23-0912839	2025-10-30	Erythrocyte Sedimentation Rate (ESR)	2025-10-30	Pacific Diagnostic Center	View
Lab Report - Shaun D. Berg.pdf	11/16/2025	Vision	Shaun D. Berg	FH-00751288	257 Clearwater Dr, Richland, WA 99352	1980-09-17	Christina O. Munroe, MD	-	2025-11-06	Cortisol, AM Serum	2025-11-06	Tri-State Endocrine Labs	View
Lab Report - Richard M. Guerrero.pdf	11/16/2025	Vision	Richard M. Guerrero	FH-0027855	1234 Medical Plaza Drive, Seattle, WA 98101	1956-08-02	Allison P. Bae, MD	ACC-23-0912213	2025-10-31	Prostate Specific Antigen (PSA)	2025-11-01	Northwest Clinical Laboratories	View
Lab Report - Patricia W. Kim.pdf	11/16/2025	Vision	Patricia W. Kim	FH-00358119	132 Harbor View Dr, Anacortes, WA 98221	1949-06-19	Louis H. Barrett, MD	-	2025-10-28	C-Reactive Protein, High Sensitivity	2025-10-29	Tri-State Endocrine Labs	View
Lab Report - Noah T. Caldwell.pdf	11/16/2025	Vision	Noah T. Caldwell	FH-00339784	104 Aspen Ridge Rd, Pullman, WA 99163	1951-10-02	Andrea G. Leung, MD	-	2025-10-27	BNP (B-Type Natriuretic Peptide)	2025-10-27	-	View
Lab Report - Monica L. Salinas.pdf	11/16/2025	Vision	Monica L. Salinas	FH-00518390	1234 Medical Plaza Drive, Seattle, WA 98101	1969-01-21	Jose E. Vargas, MD	ACC-23-0912422	2025-10-29	D-Dimer, Quantitative	2025-10-29	Pacific Diagnostic Center	View
Lab Report - Marisol Y. Delgado.pdf	11/16/2025	Vision	Marisol Y. Delgado	FH-00729184	-	1982-03-22	Jason M. Koh, MD	ACC-23-0911475	2025-10-28	Complete Blood Count (CBC)	2025-10-29	Northwest Clinical Laboratories	View
Lab Report - Lauren B. McIntyre.pdf	11/16/2025	Vision	Lauren B. McIntyre	FH-00901354	217 Lake-Shore Dr, Spokane, WA 99203	1998-07-14	Hannah R. Steele, DO	-	2025-11-02	Thyroid Stimulating Hormone (TSH)	2025-11-03	Tri-State Endocrine Labs	View
Lab Report - Keith A. Fulton.pdf	11/16/2025	Vision	Keith A. Fulton	FH-00574612	1234 Medical Plaza Drive, Seattle, WA 98101	1964-02-01	Sarah P. Lo, MD	ACC-23-0911723	2025-11-01	Hemoglobin A1c	2025-11-02	Pacific Diagnostic Center	View
Lab Report - Isabel Q. Choi.pdf	11/16/2025	Vision	Isabel Q. Choi	FH-00883862	1500 Orchard Way, Kent, WA 98032	2002-04-09	Kevin A. Dunlap, PA-C	ACC-23-0912298	2025-11-07	MonoSpot (EBV) Rapid Test	2025-11-07	Cascade Pathology Laboratory	View

IOERD Fax Processing System PoC - User Interface

Document Detail – Review, Edit, Export

Search / Lab Report - Sofia N. Morales.pdf

Lab Report - Sofia N. Morales.pdf

[Edit](#) [Export](#)

Document Information

Filename: Lab Report - Sofia N. Morales.pdf

MD5 Hash: c10c908ebf1996a3056a3490c22023f8

Processed: 2025-11-16 21:20:29.521647

Last Updated: 2025-11-16 21:33:01.994660

Processing Method:

Extracted Data

Patient Full Name
Sofia N. Morales

Patient Identifier
FH-00660322

Patient Address
2142 Highland Trail, Seattle, WA 98112

Patient Dob
02/23/1989

Ordering Provider Name
Henry J. Olson, MD

Specimen Number
ACC-23-0912839

Collection Date
10/30/2025

Test Name
Erythrocyte Sedimentation Rate (ESR)

Result Date
10/30/2025

Performing Lab
Pacific Diagnostic Center

Document Content View (raw & markdown)

[Raw Text](#) [Vision Markdown](#)

[Copy](#) [Raw](#)

Pacific Diagnostic Diagnostic Center

CLIA Certified Laboratory • CAP Accredited 1234 Medical Plaza Drive, Seattle, WA 98101 Phone: (206) 555-0100 • Fax: (206) 555-0101 lab@pacificdiagnosticcenter.com

PATIENT INFORMATION

Patient Name:	Sofia N. Morales
Date of Birth:	02/23/1989
Patient ID:	FH-00660322
Specimen Number:	ACC-23-0912839
Address:	2142 Highland Trail, Seattle, WA 98112

TEST INFORMATION

Ordering Provider:	Henry J. Olson, MD
Collection Date:	10/30/2025
Ordering Provider:	Henry J. Olson, MD

IOERD Fax Processing System PoC - API Interface

The system has an API endpoint that can be used to monitor the IOERD system, search and retrieve records, create and update records and start and top document processing. *It can also be used by generative AI systems via external tool usage or MCP.*

API Documentation

Method	Endpoint	Description
GET	/	Home
GET	/health	Health Check
GET	/health/ready	Readiness Check
GET	/health/live	Liveness Check
GET	/monitoring	Monitoring Dashboard
GET	/api/monitoring/health	Api System Health
GET	/api/monitoring/metrics	Api System Metrics
GET	/api/monitoring/errors	Api Error Reports
POST	/api/monitoring/export	Api Export Logs
GET	/api/system/info	System Info
POST	/api/monitoring/start	Start Monitoring
POST	/api/monitoring/stop	Stop Monitoring
POST	/api/monitoring/scan	Trigger Manual Scan
GET	/api/documents	Search Documents
POST	/api/documents/search	Search Documents Post
GET	/api/documents/{document_id}	Get Document
PUT	/api/documents/{document_id}	Update Document
GET	/api/documents/{document_id}/export	Export Document
GET	/document/{document_id}	View Document
GET	/document/{document_id}/edit	Edit Document
POST	/api/documents/{document_id}/reprocess	Reprocess Document
GET	/search/advanced	Advanced Search
GET	/api/search/export	Export Search Results

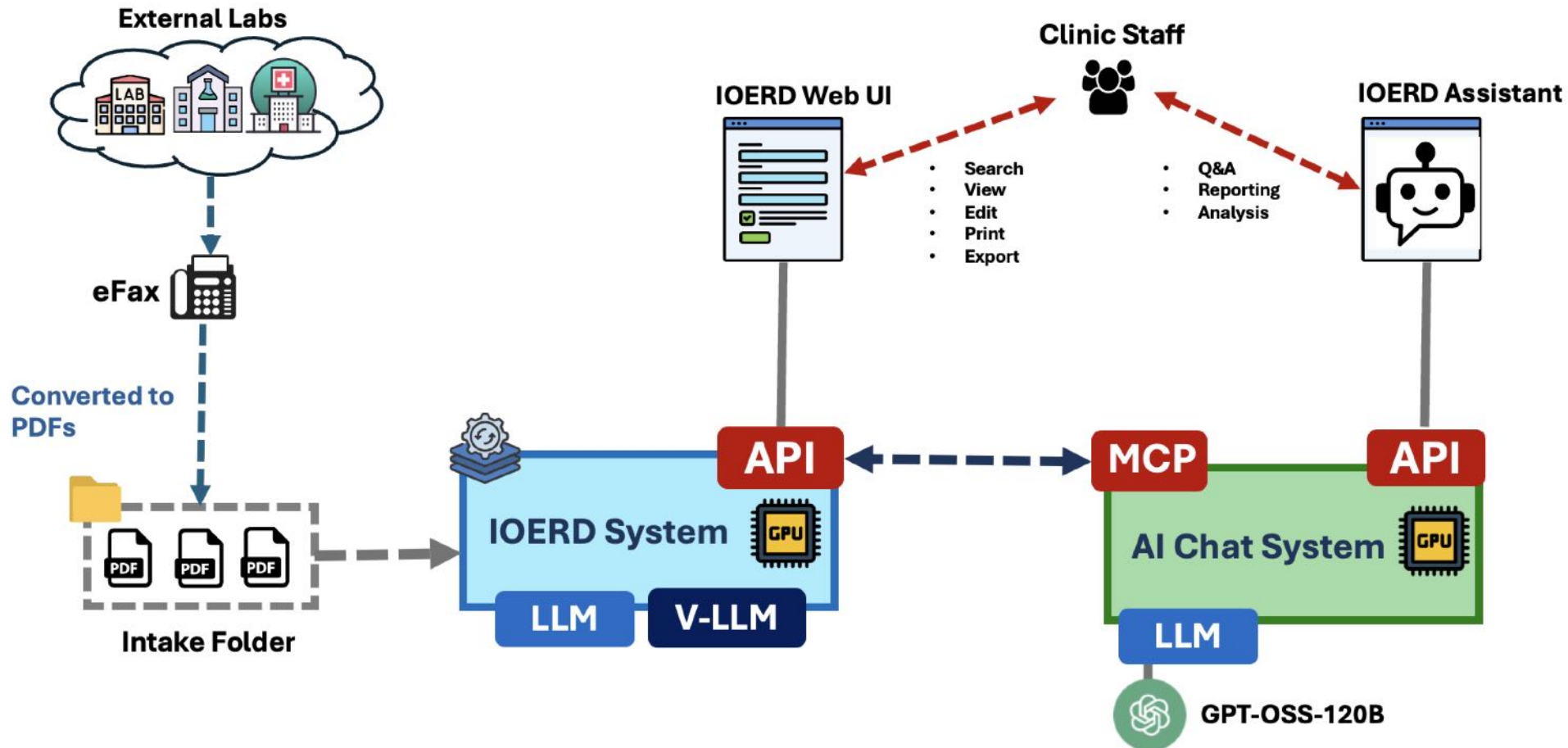
Example using the API to retrieve a record

```
curl -sS -X 'POST' \
  'http://localhost:8000/api/documents/search' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
    "field_filters": { "patient_identifier": "FH-00574612" },
    "text_search": "", "limit": 50, "offset": 0,
    "order_by": "processed_at","order_desc": true
  }' \
| jq '.documents | map({
  id, filename, patient_identifier, patient_full_name, patient_address,
  patient_dob, ordering_provider_name, specimen_number, collection_date,
  test_name, result_date, performing_lab
})'
[
  {
    "id": 11,
    "filename": "Lab Report - Keith A. Fulton.pdf",
    "patient_identifier": "FH-00574612",
    "patient_full_name": "Keith A. Fulton",
    "patient_address": "1234 Medical Plaza Drive, Seattle, WA 98101",
    "patient_dob": "1964-02-01",
    "ordering_provider_name": "Sarah P. Lo, MD",
    "specimen_number": "ACC-23-0911723",
    "collection_date": "2025-11-01",
    "test_name": "Hemoglobin A1c",
    "result_date": "2025-11-02",
    "performing_lab": "Pacific Diagnostic Diagnostic Center"
  }
]
```

IOERD Fax Processing System PoC - AI Assistant

IOERD AI system integration via API/MCP with models that support tool calling

The IOERD System's OpenAPI API interface can be used by AI systems. The LLM used must have been fine-tuned to support tool calling. For good results, a relatively large LLM should be used, such as GPT-OSS-120B. This system can be completely self-hosted (on-prem if desired), so the organization can maintain full control of the data and avoids the need for BAAs and trusting external vendors to protect our sensitive data.



IOERD Fax Processing System PoC - AI Assistant

IOERD System Chat AI Agent Examples

IOERD System Chat ▾ +

Robert McDermott Yesterday at 11:00 AM
Hello what are you, what is your purpose, and how can you help me?

IOERD System Chat Yesterday at 11:07 PM
Hello! I'm a medical system agent for the Fred Hutchinson Cancer Center, specifically designed to assist clinical operations teams.

My Purpose:
I help you search and retrieve orders that have been processed and stored in the IOERD System (medical document system).

How I Can Help You:

- **Look up medical orders** - I can search for specific orders by various criteria
- **Retrieve order information** - I can find details about processed orders
- **Search medical documents** - I can query the IOERD database to find the information you need

How to Use Me:
Simply tell me what order information you're looking for, such as:

- Searching by patient information
- Looking up specific order types
- Finding orders by date ranges
- Retrieving order details by ID or other identifiers

What would you like me to help you find today?

Robert McDermott Today at 11:10 PM
Create a formatted table with an inventory of the last 5 records that have been received





IOERD System Chat Today at 11:10 PM
Based on the records in the IOERD System, here is an inventory of the last 5 records received N/A :


RECORD ID	PATIENT NAME	PATIENT ID	TEST NAME	ORDERING PROVIDER	COLLECTION DATE	RESULT DATE	PERFORMING LAB	PROCESSED AT
20	Trevor S. Holt	FH-00601467	Serum Magnesium	Karen D. Ishikawa, MD	11/01/2025	11/01/2025	SoundLab Diagnostics	11/16/2025 21:22:18
19	Sofia N. Morales	FH-00660322	Erythrocyte Sedimentation Rate (ESR)	Henry J. Olson, MD	10/30/2025	10/30/2025	Pacific Diagnostic Center	11/16/2025 21:20:29
18	Shaun D. Berg	FH-00751288	Cortisol, AM Serum	Christina O. Munroe, MD	11/06/2025	11/06/2025	Tri-State Endocrine Labs	11/16/2025 21:18:31
17	Richard M. Guerrero	FH-0027855	Prostate Specific Antigen (PSA)	Allison P. Bae, MD	10/31/2025	11/01/2025	Northwest Clinical Laboratories	11/16/2025 21:16:42
16	Patricia W. Kim	FH-00358119	C-Reactive Protein, High Sensitivity	Louis H. Barrett, MD	10/28/2025	10/29/2025	Tri-State Endocrine Labs	11/16/2025 21:14:51

Summary: The system shows 5 of 20 total records, all processed on 11/16/2025.

IOERD Fax Processing System PoC - AI Assistant

IOERD System Chat AI Agent Examples

 **Robert McDermott** Today at 11:16 PM
A lab result was received for sample ACC-23-0912144 but it was missing the patient information and the ordering clinicians name. Do you have any information on this sample?
  

 **IOERD System Chat** Today at 11:16 PM
Yes, I found the information for specimen ACC-23-0912144 in our IOERD System

Patient Information:

- **Patient Name:** Gina R. Whitehorse
- **Patient ID:** FH-00936742
- **Date of Birth:** 10/24/1985

Ordering Clinician:





- **Ordering Provider:** Samuel T. Lin, MD


Additional Test Details:

- **Test Ordered:** Pregnancy hCG, Serum Qualitative
- **Collection Date:** 11/03/2025
- **Result Date:** 11/03/2025
- **Performing Lab:** Pacific Diagnostic Center
- **Result:** hCG <5 mIU/mL (Negative)

The complete lab report is stored in our system under filename "Lab Report - Gina R. Whitehorse.pdf" and was processed on 11/16/2025.

Is there any additional information you need about this specimen?

 **Robert McDermott** Today at 11:21 PM
What are the lab results for patient "FH-00729184"?
  

 **IOERD System Chat** Today at 11:21 PM
Based on the records in the IOERD System, here are the lab results for patient FH-00729184:

Patient Information

- **Patient Name:** Marisol Y. Delgado
- **Date of Birth:** 03/22/1982
- **Patient ID:** FH-00729184
- **Specimen Number:** ACC-23-0911475

Test Information

- **Test Ordered:** Complete Blood Count (CBC)
- **Ordering Provider:** Jason M. Koh, MD
- **Collection Date:** 10/28/2025
- **Result Date:** 10/29/2025
- **Performing Lab:** Northwest Clinical Laboratories


Laboratory Results


TEST NAME	RESULT	UNIT	REFERENCE RANGE	STATUS	⌵
WBC	5.9	×10 ³ /μL	4.0–11.0	Normal	
RBC	4.56	×10 ⁶ /μL	4.20–5.90	Normal	
Hemoglobin	13.8	g/dL	12.0–16.0	Normal	
Hematocrit	41	%	36–46	Normal	
Platelets	243	×10 ³ /μL	150–450	Normal	

Report Summary: All CBC values are within normal reference ranges. The report was electronically signed by Dr. James R. Patterson, MD, PhD on 10/29/2025 at 14:32 PST.

IOERD Fax Processing System PoC - AI Assistant

IOERD System Chat AI Agent Examples

 **Robert McDermott** Today at 11:26 PM
Generate some type of graphic to visualize the data currently in the IOERD system

 **IOERD System Chat** Today at 11:26 PM

IOERD System Data Visualization

Based on the current data in the IOERD system, here's a visual representation:

System Overview

Total Documents: 20 records

Data Completeness by Field

Field	Fill Rate (% Complete)
Patient Name	100%
Patient ID	100%
Patient DOB	100%
Ordering Provider	100%
Collection Date	100%
Test Name	100%
Result Date	100%
Patient Address	95%
Performing Lab	90%
Specimen Number	45%

Data Distribution Summary

METRIC	COUNT	NOTES
Unique Patients	20	All unique patient records
Unique Providers	20	20 different ordering clinicians
Unique Tests	20	Variety of test types
Performing Labs	6	Tests distributed across 6 facilities
Collection Dates	13	Spanning multiple collection days

Data Quality Alert

Specimen Number field shows only 45% completion (9 out of 20 records). This may require attention for complete traceability.

Recent Activity

All 20 records were processed on **November 16, 2025**

Would you like me to provide more detailed analysis of any specific aspect of this data?

Thank You 🙌